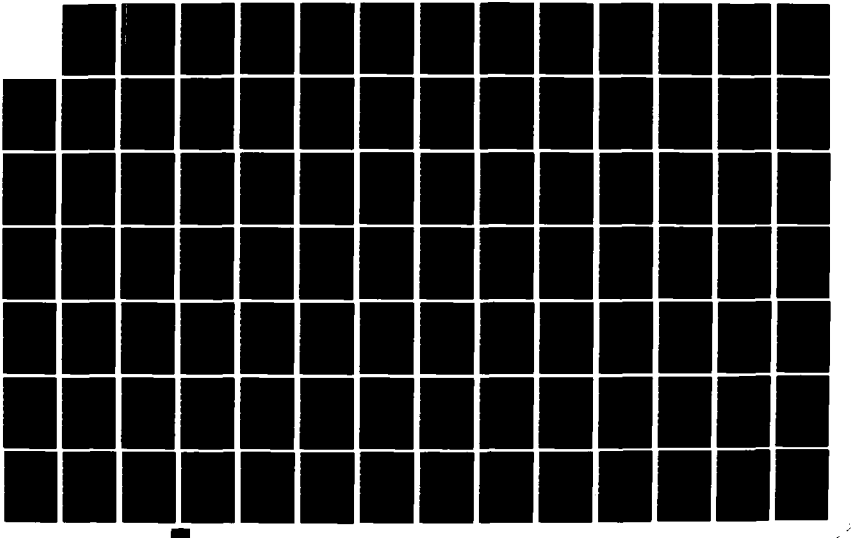
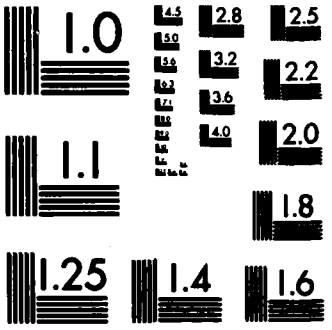
AD-A133 554 ADVANCED TELEPROCESSING SYSTEMS(U) CALIFORNIA UNIV LOS ANGELES DEPT OF COMPUTER SCIENCE L KLEINROCK 30 SEP 82 UCLA-ENG-83-24 MDA903-82-C-0064 1/2

UNCLASSIFIED F/G 17/2.1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

ADVANCED TELEPROCESSING SYSTEMS
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

ANNUAL TECHNICAL REPORT

SEPTEMBER 30, 1982

Principal Investigator: Leonard Kleinrock

Computer Science Department
School of Engineering and Applied Science
University of California
Los Angeles

UNIVERSITY OF CALIFORNIA, LOS ANGELES

SEMI-ANNUAL TECHNICAL REPORTS

Sponsored by the

Defense Advanced Research Projects Agency

Contract Nos.: DAHC-15-C-0368, MDA 903-77-C-0272; DARPA Order No. 2496

COMPUTER NETWORK RESEARCH

DATES	DDC ACCESSION NUMBER
August 1969 to February 1970:	AD 705 149
August 1970:	AD 711 342
June 1971:	AD 727 989
December 1971:	AD 739 705
June 1972:	AD 746 509
December 1972:	AD 756 708
June 1973:	AD 769 706
December 1973:	AD A004167
June 1974:	AD A008422
December 1974:	AD A016823
June 1975:	AD A020671
December 1975:	AD A025914
June 1976:	AD A034171 (Final)

ADVANCED TELEPROCESSING SYSTEMS

June 1976 to December 1976:	AD A039018
June 1977:	AD A047496
June 1978:	AD A077404
September 1979:	AD A081938
March 1980:	AD A088839
September 1981:	to be assigned

ADVANCED TELEPROCESSING SYSTEMS

Annual Technical Report

September 30, 1982

Contract Number: MDA 903-82-C-0064
DARPA Order Number: 2496
Contract Period: October 1, 1981 to September 30, 1983
Report Period: October 1, 1981 to September 30, 1982

Principal Investigator: Leonard Kleinrock

Co-Principal Investigator: Mario Gerla

(213) 825-2543

Computer Science Department
School of Engineering and Applied Science
University of California, Los Angeles

Sponsored by

DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: UCLA-ENG-83-24	2. GOVT ACCESSION NO.: AD-A133554
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): ADVANCED TELEPROCESSING SYSTEMS: ANNUAL TECHNICAL REPORT	5. TYPE OF REPORT & PERIOD COVERED: Annual Technical 1 OCT 81 through 30 SEPT 82
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Leonard Kleinrock	8. CONTRACT OR GRANT NUMBER(s): MDA 903-82-C-0064
9. PERFORMING ORGANIZATION NAME AND ADDRESS: School of Engineering and Appoied Science University of California Los Angeles, CA 90024	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: DARPA Order No. 2496
11. CONTROLLING OFFICE NAME AND ADDRESS: Defense Advanced Research Projects Agency 1400 Wilson Boulevard Arlington, VA 22209	12. REPORT DATE: September 30, 1982
	13. NUMBER OF PAGES: 151
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office):	15. SECURITY CLASS. (of this report):
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Computer communication networks, Packet switching, Multi-hop networks, Packet radio networks, Channel access Protocols, ALOHA network, Spatial-TDMA, rude-CSMA, Resource sharing, Multi-access communications, Computer communications, Broadcast communications, Channel sharing.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This Annual Technical Report covers research carried out by the Advanced Teleprocessing Systems Group at UCLA under DARPA Contract No. MDA 903-82-C-0064 covering the period from October 1, 1981 to September 30, 1982.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(20)

This contract has three primary designated research areas: packet radio systems, resource sharing and allocation, and distributed processing and control.

This report contains the abstracts of the publications which summarize our research results in those areas during this annual period, followed by the main body of the report which is devoted to a study of channel access protocols that are executed by the nodes of a network to schedule their transmissions on multi-access broadcast channel. In particular, the main body consists of the Ph.D. dissertation by Randolph Nelson, "Channel Access Protocols for Multi-Hop Broadcast Packet Radio Networks," conducted under the supervision of Professor Leonard Kleinrock (Principal Investigator for this research). This work discusses some new channel access protocols useful for mobile radio networks. Included is an analysis of slotted ALOHA and some tight bounds on the performance of all possible protocols in a mobile environment. A new protocol known as Spatial TDMA is also introduced and analyzed; it is collision-free protocol which allows one to find the minimum possible delay for messages in a network which is perfectly scheduled. Yet another protocol known as Rude CSMA is introduced and studied. Here a terminal is permitted to transmit even when he detects another terminal's transmission, under certain circumstances which are discussed in detail; the cases where such an access protocol makes sense are discussed.

ADVANCED TELEPROCESSING SYSTEMS

Defense Advanced Research Projects Agency
Annual Technical Report

September 30, 1982

INTRODUCTION

This Annual Technical Report covers research carried out by the Advanced Teleprocessing Systems Group at UCLA under DARPA Contract No. MDA 903-82-C-0064 covering the period from October 1, 1981 through September 30, 1982. Under this contract we have three designated tasks as follows:

TASK I. PACKET RADIO SYSTEMS

The extension of our analytic and design techniques to modern multi-hop packet radio networks will be studied. The applications and extensions include access methods, large network control and management, queueing network models, approximation methods, capture phenomena, conflict-free algorithms, reliability, routing procedures, topological studies, TDMA in a multi-hop environment, and multiplexing methods.

TASK II. RESOURCE SHARING AND ALLOCATION

Extended concepts of "power" in networks will be studied. The extensions include more complex topologies and configurations, extended queueing disciplines, general distributions, other definitions of power, effects of varying the traffic matrix, fairness. The problems of large scale internetting with respect to resource allocation and sharing will also be studied further.

TASK III. DISTRIBUTED PROCESSING AND CONTROL

Overall principles of distributed processing and distributed control will be studied. The issues of sequencing in data base updates, distributed control, and distributed processing (involving the calculation of concurrency of processing) are the subjects of concern here.

A major contribution of our research during this period is contained in Reference 5 listed below, namely "Channel Access Protocols for Multi-Hop Broadcast Packet Radio Networks", by Randolph Nelson. This Ph.D. dissertation (supervisored by Professor Leonard Kleinrock) makes a number of contributions to the field of access protocols for multi-hop systems. In this report, slotted ALOHA is analyzed in a mobile packet radio network. Furthermore, a new protocol known as *Spatial-TDMA* is described and analyzed; this is a collision-free system for which we find the mean response time and allows us to determine the absolute minimum delay averaged over all messages in the network. Yet

another protocol referred to as *Rude-CSMA* is discussed and analyzed and is an access protocol which permits terminals to transmit even after they sense a busy channel. The performance of the system and the conditions under which it is useful are discussed. The entire dissertation is reproduced as the main body of this report immediately following the list of research publications below. This list of publications summarizes the results of this annual period and the abstract of each paper is given along with the reference itself.

RESEARCH PUBLICATIONS

1. **Nelson, R. and L. Kleinrock, "Spatial TDMA: A Collision Free Multi-Hop Channel Access Protocol,"** *ICC '82 Conference Proceedings,* **Philadelphia, Pennsylvania, June 13-17, 1982, pp. 1C.4.1 - 1C.4.4, submitted to** *IEEE Transactions on Communications.*

   In this paper we define a broadcast channel access protocol called *Spatial-TDMA* which is designed specifically to operate in a multi-hop packet radio environment where nodes are assumed to be stationary. The protocol defined assigns transmission rights to nodes in the network in a local TDMA fashion and is collision-free. Methods for determining slot allocations and slot sizes are developed and a solution is found for determining the assignment of capacities for the links of the network that minimizes the average delay of messages in the system.

2. **Levy, H., "Record Recognition and Duplicate Elimination of Personnel Lists," Computer Science Department, University of California, Los Angeles, Master's Thesis, June 1982.**

   Personnel lists consist of records which contain descriptions of individuals. Each such description can be the person's name, address, birth date, profession, etc. Due to misspelling or other reasons these descriptions are not *mathematically exact* in their computer description. For this reason, it is very likely that one person can be represented in different lists by records which are not exactly identical to each other.

   In the presence of this property it is desired to *recognize* personnel records. More specifically, it is required to identify if two records r_1, r_2 represent the same person.

   This problem is confronted in this work. While the solutions offered by existing methods are based on heuristics only, we propose a method which combines heuristics with the statistical behavior of the lists and which finds a suboptimal decision criteria for this problem. Assuming that a *similarity* measure is defined, we find an optimal selection method (which identifies if two records are *the same*), based on this measure.

3. **Takagi, H. and L. Kleinrock, "Output Processes in Contention Packet Broadcasting Systems," Computer Science Department Report No. CSD-821006, July 1982.**

The processes consisting of the packet interdeparture times in contention-type packet broadcasting systems are studied. The channel access protocols considered include slotted and unslotted ALOHA and carrier-sense-multiple-access (CSMA) with collision detection or with delay capture effect. Through analysis of the channel activity cycle, the distribution, mean and coefficient of variation of the packet interdeparture times are explicitly derived. Taking the reciprocal of the mean interdeparture time, we obtain the channel throughput. Cases with dissimilar users are also considered. Application of the present results to the packet queueing processes is suggested.

4. **Kleinrock, L., "A Decade of Network Development,"** *Journal of Telecommunications Network,* **Spring 1982, pp. 1-11.**

The network technologies developed over the past decade, due largely to incredible advances in integrated chip production, have set the stage for an enormous DP revolution the impact of which will soon be felt in the business community and in the home and consumer markets. We have been able to identify those fundamental principles which render this new technology so cost effective. End-to-end communication services are now coming into place which will pass data, voice, video, fax, graphics, etc. from an individual terminal in an office or a living room across vast distances to remote processing facilities. The technological issues have been addressed, and in this paper we describe some of the successes and some of the remaining problems to be solved. As engineers, we cannot restrict our efforts to the purely technical (and "nice") problems, but rather we must accept the further responsibility of familiarizing ourselves with the overall environment in which our "products" must perform; if we fail in this regard, the world of commerce, trade and business will ignore our best efforts as irrelevant to its needs.

5. **Nelson, R., "Channel Access Protocols for Multi-Hop Broadcast Packet Radio Networks," Ph.D. Dissertation, Computer Science Department, University of California, Los Angeles, July 1982, Report No. CSD-820731.**

UCLA-ENG-82-59
July 1982

CHANNEL ACCESS PROTOCOLS FOR MULTI-HOP BROADCAST PACKET RADIO NETWORKS

by
Randolph Nelson

This research, conducted under the chairmanship of Professor Leonard Kleinrock, was sponsored by the Defense Advanced Research Projects Agency, Department of Defense.

Computer Science Department
School of Engineering and Applied Science
University of California
Los Angeles

ABSTRACT

A multi-hop packet radio network consists of a set of geographically distributed nodes (e.g. computers, terminals, etc.) which communicate using radio transceivers over a shared, broadcast radio channel. In this dissertation we study algorithms, called *channel access protocols*, that are executed by the nodes of the network to schedule their transmissions on the channel. In particular, we analyze the performance obtained in mobile networks where nodes use the slotted-ALOHA protocol and determine network parameters that maximize throughput. These results are extended when we obtain a tight upper bound for the performance of all protocols in a mobile environment. This bound can be used as a standard by which we compare the performance of other protocols in multi-hop mobile networks and we show that slotted-ALOHA is about 36% efficient in such environments.

In a related area, we define a new protocol, *spatial-TDMA*, designed expressly for networks in which nodes are stationary. This protocol is collision-free and we find the mean delay for messages passing through nodes using the protocol. This leads us to formulate and solve a capacity assignment problem for networks using spatial-TDMA, which minimizes the delay averaged over all messages in the network.

Another protocol we define, *rude-CSMA*, uses information only about a node's local environment. This protocol is called "rude" because, under certain circumstances, the throughput of the network is increased if nodes, after sensing a busy channel, transmit packets with a non-zero rate. We derive the performance equations for this protocol and then find system parameters that maximize throughput for different network topologies.

ACKNOWLEDGEMENTS

I would like to express my appreciation to my doctoral committee consisting of Professors Leonard Kleinrock, Mario Gerla, James Jackson, Richard Muntz, and R. Clay Sprowls. Special thanks goes to Dr. Leonard Kleinrock, my dissertation chairman, for his encouragement and knowledgeable guidance during the course of this work. It was sheer joy working with someone so enthusiastic and insightful.

The support by the Advanced Research Projects Agency of the Department of Defense contract number MDA 903-77-C-0272 for this work is greatly appreciated and I owe a debt of gratitude to the many employees of that contract. I would especially like to thank Trudy Cook, Amanda Daniels, George Ann Hornor, Linda Infeld, Lillian Larijani, and Terry Peters for their invaluable aid and friendship during my stay at UCLA. We can all thank Ruth Pordy for the exemplary art work found within the pages of the dissertation.

I would like to thank Mart Molle for so generously helping me become familiar with the local computer system and to Richard Gail for helping me circumvent the numerous bugs and quirks of the system! To the other students of the ATS group, Abdelfettah Belghith, Rina Dechter, Joe Green, Ken Kung, Hanoch Levy, and Hideaki Takagi, I offer my sincere thanks for making the environment stimulating and exciting. Students outside the ATS group who have added immeasurably to my experience include Joe Hellerstein, Paul Hurley, Debbie Lewis, Mike Molloy, and Mary Vernon. It has been my very great pleasure and honor to become close friends with Hamid Pirahesh and I appreciate all the help he has given me throughout these years.

To Cindy, to whom I dedicate this dissertation, I owe everything. Life would not be the same without her understanding and playfulness. Her love was a necessary and sufficient condition for my happiness.

The author would be seriously amiss if he did not acknowledge Peter Denning whose wise counsel made the writing of this acknowledgement page (let alone this dissertation!) possible.

Table of Contents

List of Figures

LIST OF NOTATION

The following is a list of notations used in the thesis. Notations are consistent throughout and we have organized the list according to the chapter in which the notation first appears.

CHAPTER 2

λ	Mean density of terminals on the plane.
R	Transmission radius of nodes in the network.
N	The average number of neighbors.
β	The capture parameter of the network.
n	The number of nodes in the network.
p	The probability of transmitting.
N'	The effective number of neighbors.
E	The expected number of successes in the network.
G	The offered load of packets from a circle of radius R.
G'	The effective offered load.

CHAPTER 3

R_0	The radius of the entire network.
R_1	The random variable for the euclidian distance for nodes one hop away.
R_2	The random variable for the euclidian distance for nodes two hops away.
R_3	The random variable for the euclidian distance for nodes three hops away.
Θ	The random variable for the angle between the vectors of the first and second hop.
$f^*(N)$	The maximum fraction of successful transmissions.

CHAPTER 4

T	The size of the time frame.
C_i	Clique i.
t_i	The total time from the time frame allocated to C_i.
$\underline{t}$	The time vector.
$\|\|\underline{t}\|\|$	The sum of the elements of $\underline{t}$, $\|\|\underline{t}\|\| = \sum_{i=1}^{k} t_i$.
C	The set of all cliques, $C = \{C_1, C_2, \cdots, C_k\}$.
$\gamma_{i,j}$	The average flow over a time frame of messages from node i to node j.
γ	The total flow of messages in the network over a time frame.
α_i	The average flow of messages over a time frame over arc i.
T_{ex}	The external arrival time allocated to a specified queue over one time frame.
T_{in}	The internal arrival time allocated to a specified queue over one time frame.
T_s	The service time allocated to a specified queue over one time frame.
L_{ex}	The average number of external arrivals to a specified queue over one time frame.
L_{in}	The average number of internal arrivals to a specified queue over one time frame.
M_{ex}	The rate used in the fluid approximation over external arrival periods for a specified queue.
M_{in}	The rate used in the fluid approximation over internal arrival periods for a specified queue.
M_s	The rate used in the fluid approximation over service periods for a specified queue.
T_{ex}^i	T_{ex} corresponding to queue i.
T_{in}^i	T_{in} corresponding to queue i.
T_s^i	T_s corresponding to queue i.
L_{ex}^i	L_{ex} corresponding to queue i.
L_{in}^i	L_{in} corresponding to queue i.
$D_i(\underline{t})$	The average delay of messages passing through queue i for time vector $\underline{t}$.

CHAPTER 5

x	Parameter to rude-CSMA protocol.
y	Parameter to rude-CSMA protocol.
γ_0	Average rate of message arrival to nodes in rude-CSMA.
μ	Average transmission time.
ρ	Equal to γ_0/μ.
$s_i(S)$	The state of node i, $s_i(S)=1$ means node i is transmitting, otherwise it is not transmitting.
S	The state of the system, $S=(s_1(S), s_2(S), \cdots, s_n(S))$.
$N_1^i(S)$	The number of neighbors of node i, in state S, that are transmitters.
$N_0^i(S)$	The number of neighbors of node i, in state S, that are not transmitters.
$B_1(S)$	The number of adjacent transmitters in state S.
$B_0(S)$	The number of adjacent non-transmitters in state S.
$M(S)$	The number of transmitters in state S.
$E(S)$	The expected amount of successful transmissions in state S.
$r_0^i(S)$	The rate at which node i presents packets to the channel in state S.
$r_1^i(S)$	The rate at which node i ends its transmission in state S.
$\Pi(S,x,y)$	The probability of being in state S for parameter values x and y.
$E(x,y)$	The expected throughput for parameter values x and y.
N^*	The effective number of idle neighbors.

CHAPTER 1
INTRODUCTION

In this section we will briefly recount some of the events that lead to the development of computer networks. We will then define terms used to describe packet radio networks and delineate the problems which are addressed in this dissertation.

1.1 A Little History

The basic problems of computer science are those involved with sharing resources. In the 1960's, since computing machinery was expensive, designers of computer systems broke away from the one-job one-machine system configuration and developed the policy of multiprogramming. This decreased the cost of the machine time dedicated to each job, but required considerable effort to create and implement operating system software. As the complexity of the system environment increased, so did the intricacies of the operating system, and issues concerning its functionality and implementation became major efforts. In some sense the operating system became another component of the resource sharing problem it was designed to solve. The difficulty associated with managing the complexity of these systems, combined with economics, served as a catalyst to steer things away from building larger systems. Mass fabrication of integrated circuitry drastically reduced the cost of computers to the point where it was no longer vital to operate systems at high utilizations to be cost effective, and thus providing more computers, instead of making them larger and more powerful, became more feasible. The resource sharing problems that were created with this trend consisted of developing efficient methods to allow distributed users to share many smaller machines rather than providing them sufficient processing power from one large, commonly shared machine. Besides problems in distributing processing, this approach created the need to connect these machines with a communications network. The resource sharing problems of operating systems were thus extended to include those associated with computer networking.

Experimental networks, such as the ARPANET of the department of defense [Klei76a], demonstrated the feasibility of the network approach. This network used point to point cable to link the communication processors of the network together, which, since the network itself was a shared resource, became the next part of the resource sharing problem. The inflexibility of point to point cable to changes in the topology of the network motivated network designers to seek a more flexible transmission media for the links of the network. The successful implementation of the ALOHA-NET [Abra70a] of the University of Hawaii showed that it was possible to use broadcast radio as this new media. A new resource then, the radio

spectrum, was added to the set of resources which had to be efficiently shared.

In such networks, if two or more nodes transmit simultaneously, their signals *collide* and neither signal is successfully received. Thus the resource sharing problem of scheduling transmissions on the channel to minimize the number of collisions in the network became an important research area. This brief historical account leads us to the present day and to the topic of this dissertation.

Our major purpose in this dissertation is to lend understanding to these scheduling algorithms, called *channel access protocols*, for radio networks. Before delineating the set of problems addressed in this dissertation and its major contributions, we must make a few definitions in the next section. In the sections following this, we will then relate in a more precise manner, the basic principles of resource sharing with the results of our investigations.

1.2 Preliminary Definitions

In this section we briefly define the essential elements of a *packet radio network*. A *packet radio unit* is a hardware device consisting of a *communications section* which transmits and receives signals, and a *logic section* which constructs, decodes, and processes messages. The unit of information sent by the transmitter is called a *packet*. We will distinguish two types of radio units, *regular packet radios* and *repeaters*. This classification depends on the nature of the programs the unit executes and we allow units to change from one classification to the other. Typically the function of repeaters is to form communication links between regular units which are separated by distances greater than their transmission range, and thus relay messages from one unit to another. To prevent repetitious language, we will use *terminals* and *nodes* to mean regular packet radio units.

The *topology* of the network is the spatial placement and transmission range of the nodes of the network. The *channel* is the communications media used by the transceivers and in our work will be broadcast radio. We say the channel is *idle* with respect to a certain node at a certain point in time, if that unit does not detect a signal on the channel at that time, and that the channel is *busy* otherwise.

Suppose a terminal t_1 is sending a message to terminal a and that a is within range of t_1's transmission. Also assume that t_1's message is sent over the time period $[0,t]$. We say that a captures t_1's signal at time τ, $(0 \leqslant \tau \leqslant t)$, if a can decipher the information in t_1's signal at τ. We say that t_1's transmission is *successful* if a captures t_1's signal over all τ in $[0,t]$ otherwise we say t_1's signal *collided* with other signals (possibly noise) on the channel. In a typical radio environment there are two parameters that govern the capture of signals, power and time. In *power-capture*, t_1's signal will be received by a in the presence of a transmission by t_2, if the ratio of the power of the two signals received by a is greater than a certain number, β^{-1}, called the *capture-ratio*. If $\beta=0$ *(non-capture)* a will receive t_1's transmission only if there are no

other transmitters within a's hearing range, and if $\beta = 1.0$ *(perfect-capture)* a will receive if t_1 is closer to a than t_2. With *time-capture*, t_1's signal is received by a if t_1 began transmitting at least a certain time, the *capture time*, prior to the beginning of t_2's transmission.

Packets originating in one node may need to be sent to several intermediary nodes to reach their final destinations. We call each transmission in this relay a *hop* and distinguish between *single-hop* and *multi-hop* environments. A single-hop environment is said to be *line-of-sight* if each node detects transmissions from all the nodes in the network, and said to be *hidden* if some nodes do not hear all the nodes in the network. In the hidden environment we assume terminals can only send packets to terminals they can hear. In the multi-hop environment terminals send packets to destinations that are outside their transmission ranges and a *routing algorithm* is needed to specify the next node that is to be the recipient of the packet.

Nodes execute an algorithm called the *channel access protocol* to determine when to transmit their packets and the information used as input to this algorithm is called the *policy state information.* If terminals use data containing information about a proper subset of the nodes in the network, the policy is said to be a *local* policy, otherwise we say it is a *global* policy.

We will judge the performance of the network by its delay-throughput characteristics. The *throughput* of the network is the average number of messages reaching their final destinations over a unit time interval over a specified region of the network. In single-hop environments, this region is defined to be the entire network and thus the throughput of the network is equal to the average number of successful messages on the channel per unit time. In multi-hop environments however, several successes can occur simultaneously in different parts of the network, and thus we calculate throughput over a region of unit area. The maximum throughput of a particular network for a particular channel access protocol is called the *capacity* of the network. The *delay* of the network is the average time between message arrival at the originating node and delivery to the final node. A certain policy will be said to be *optimal* if we can show that no other policy using a subset of the policy state information achieves a greater throughput.

1.3 Statement of the Problem

The problem we will study in this dissertation consists of creating efficient transmission policies for multi-hop packet radio broadcast networks. Thus the shared resource is the broadcast radio channel and the objective is to maximize the capacity of the channel over given sets of policy state information. To elucidate some of the characteristics of this problem, let us first discuss problems in developing protocols for the single-hop line-of-sight environment. In networks of this type all nodes share identical information about the status of the channel. This information is obtained by sensing the channel and is exploited, for example, in the *CSMA* family of protocols [Toba74a]. In CSMA, a terminal wishing to transmit a packet on the channel, first senses the channel to determine if it is idle or busy. If idle, the terminal transmits its packet immediately, otherwise it refrains until some future time and again senses the channel. Collisions occur only if two or more terminals sense the channel within a *propagation delay* (defined to be the average delay between the time a message is transmitted on the channel and the recognition that this terminal is transmitting by all the nodes in the network) of each other, and hearing it idle, all transmit their packets. If the propagation delay between terminals is small, the probability of this event is also small and collisions are infrequent.

What other information would be valuable in determining efficient transmission policies for single-hop systems? Suppose the identities of the busy terminals were known by every terminal in the network. An optimal policy which both maximizes throughput and minimizes delay could then be created by giving busy terminals permission to transmit in some specified order, say first-come first-served. It is well known that the performance of such classical queueing systems are a lower bound for all possible access schemes. Such state information must be sent over the channel, though, at a cost in the delay-throughput performance of the system. Indeed this communication and the optimal use of specific control information lie at the root of the difficulty of determining optimal policies for single-hop line-of-sight environments [Moll81a]. Thus we have that to efficiently share the channel, one needs ways to distribute control information and since the channel is the only means to perform this communications, the control information requires efficient methods to use the channel. We are reminded of such a cycle mentioned in the introduction in the case of operating systems, where part of the structure it needed to manage to effectively share computing resources, was the operating system itself.

The complexity of the problem is greatly increased when the hidden environment is considered since each node hears only a proper subset of the nodes in the network, and neighboring nodes do not necessarily agree on the state of the channel. If we try transmitting to another node in the network after sensing the channel idle, as in CSMA, there is no guarantee that the receiver, who hears a different subset of nodes than the transmitter, will also sense an idle channel. The set of subnetworks created by the local information, in essence one for each terminal in the network, have a high degree of interdependence. For any two nodes, there is a set of terminals that can hear both of them, and simultaneous transmission to this *interference*

set will result in a collision.

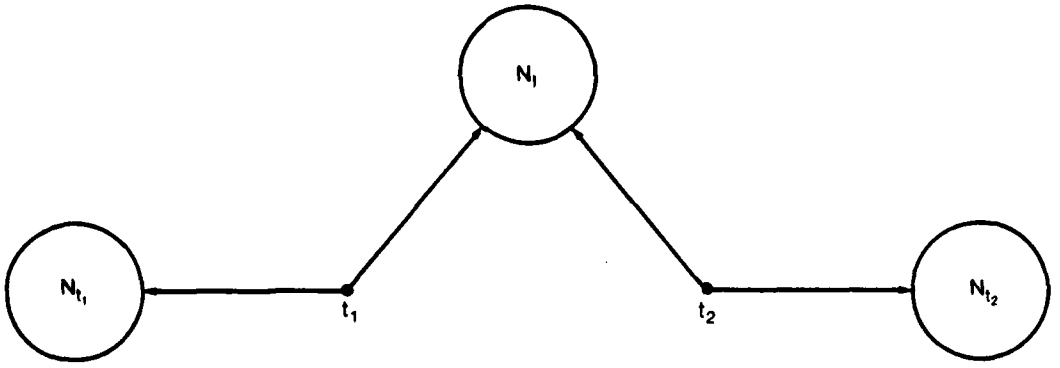


FIGURE 1.1
A Sample Network.

For example, in Figure 1.1 transmitters t_1 and t_2 have interference set N_I to which they both transmit, and sets N_{t_1} and N_{t_2} which hear only their respective transmitters. There are three distinguishable cases for successful transmission on the channel. There are no successes if both t_1 and t_2 send to N_I or if both do not transmit. One success will occur if exactly one of the terminals t_i transmits or if both transmit and one terminal sends to N_{t_i} while the other sends to N_I. Finally there are two successes if both terminals send to N_{t_1} and N_{t_2}. Coordinating their transmissions to maximize the number of successes, is called the *interference problem* and will be a major area of concern in this dissertation.

In the hidden terminal environment knowing the identities of busy terminals is not sufficient, as in one-hop networks, to generate optimal transmission policies. One must now also know the destinations for packets in each of the queues, and the traffic characteristics for future arrivals of messages.

There are some theoretical results we can prove for optimal transmission policies. In figure 1.1 let q_{t_1} and q_{t_2} be the conditional probabilities that when the transmitters transmit they send packets to N_{t_1} and N_{t_2} respectively and suppose transmitters send packets with probabilities P_{t_1} and P_{t_2}. Using the above definitions we can derive an expression for the expected number of successes. There will be one success if exactly one of t_1, and t_2 transmits, or if both transmit, if one of them transmits to the interference set (and its transmission suffers a

collision) and the other to either set N_{t_1} or N_{t_2}. Writing this out using the above notation we have:

$$P[\text{ One Success}] = P_{t_1}(1-P_{t_2}) + P_{t_2}(1-P_{t_1}) + P_{t_1}P_{t_2}((1-q_{t_1})q_{t_2} + (1-q_{t_2})q_{t_1})$$

Two successes can only occur if both transmitters send packets to sets N_{t_1} and N_{t_2}. The probability for this is:

$$P[\text{ Two Successes}] = P_{t_1}P_{t_2}q_{t_1}q_{t_2}$$

The expected number of successes is then given by:

$$E = P[\text{ One Success}] + 2P[\text{ Two Successes}]$$

$$E = P_{t_1} + P_{t_2} - P_{t_1}P_{t_2}(q_{t_1} + q_{t_2})$$

We wish to maximize this over choices of (P_{t_1}, P_{t_2}). The solution will occur in the interior of the feasible set $0 \leq P_{t_1}, P_{t_2} \leq 1$ only if the Hessian matrix (the matrix of second partial derivatives i.e. $\left(\frac{\partial^2 E}{\partial P_{t_i} \partial P_{t_j}}\right)$) is negative semi-definite [Avri76a]. The Hessian however is

$$H = \begin{pmatrix} 0 & q_{t_1} + q_{t_2} \\ q_{t_1} + q_{t_2} & 0 \end{pmatrix}$$

which is clearly not negative semi-definite since the determinant of the second principle minor is negative. We thus conclude that the optimal solution occurs at one of the vertices of the feasible set. There are four such vertices and we can eliminate (0,0) immediately since it has a throughput value of zero. This leaves three pairs to consider namely (0,1), (1,0), and (1,1). The expected number of successes for vertex (1,1) is equal to $2-(q_{t_1}+q_{t_2})$ and equals 1 for (1,0) or (0,1). We see then if $q_{t_1}+q_{t_2} \leq 1$ the optimal policy is to be rude and transmit packets regardless of the state of the channel, and if $q_{t_1}+q_{t_2} > 1$ the optimal policy lies on either the (1,0) or (0,1) vertice. We should note that to make the policy *fair*, in the sense that each user obtains an equal share of the capacity of the channel, we would alternate, in some fashion, between (1,0) and (0,1). This result generalizes to the case of n interfering terminals where the optimal policy lies on one of the vertices of a $(0,1)^n$ hypercube. To achieve maximal throughput, it does not make sense to assign transmission permission to a node if its transmission will cause a collision, and thus the above example implies that an optimal assignment of transmission permissions does not allow collisions in the network. This motivates why we define a collision-free protocol for multi-hop networks called spatial-TDMA in Chapter 4. Another observation we can make from the above example, which motivates another protocol, rude-CSMA, which we define in Chapter 5, arises from the fact that under certain circumstances nodes in the above example should transmit even if they sense a busy channel. In rude-CSMA nodes adjust the rate at which they present packets to the channel according to the

activity of the channel in their local environments and do not necessarily refrain from transmitting if they sense a busy channel. The idea, as in the example, is that a message may be destined towards a node outside the range of the other transmitters in the local environment and thus transmitting will tend to increase the throughput of the network.

We can refine the above model by first noting in the above strategy that collisions do sometimes occur, and we have assumed that terminals have packets to send at all times. If we relax these assumptions we can create a policy that maximizes the expected throughput over all time by analyzing a more complex model. Suppose now we know the messages and their destinations for all the terminal's queues as well as the topology of the network. Also suppose we know the arrival rate of new messages to each terminal and the distribution of message destinations. Then by modeling the system as a Markov decision process [Mine70a] we can assign permissions to transmit to nodes in the network and using a dynamic programming technique called policy-iteration [Howa60a] can determine a global transmission policy that will maximize the expected number of successes over all time. This decision policy is collision-free and is easily implemented if information about the queues of the network were known to all nodes. In general however, dissemination of this information would require considerable channel capacity and thus tends to counteract the optimality of the proposed policy.

1.4 Previous Work in the Area

We briefly give in this section an overview of related research in the area of channel access protocols and will concentrate on broadcast protocols for the single-hop and multi-hop environments.

Single-hop - There is a vast literature for the single-hop line-of-sight environment and we will only discuss results relevant to later sections. The most elementary contention protocol for this environment was developed by Abramson [Abra70a] and is called the *ALOHA* protocol. In this protocol a terminal transmits a message as soon as it arrives, and messages transmitted concurrently, collide and are not received by their intented nodes. Collided packets are retransmitted after nodes wait a randomly selected period of time. It can be shown that under the assumptions of an infinite population of terminals offering Poisson traffic at a constant rate, the maximum throughput of the system over a finite period of time is $1/2e$, or about 18% of the bandwidth of the channel, and that the system will eventually form an infinite backlog of messages waiting to be sent over the channel [Lam75a]. The throughput of the system can be doubled [Robe72a] by using the ingenious device of dividing the time axis into slots equal to the length of packets. This protocol, called *slotted-ALOHA*, operates in a similar fashion to ALOHA except that nodes are allowed to begin transmission only at the beginning of time

slots. This system is also unstable [Lam75a], but several methods for stablizing it have been proposed.

Observe that in the ALOHA system radio units transmit even if the channel just prior to their transmission is busy, hence guaranteeing a collision. Tobagi [Toba74a] analyzed a class of protocols called *Carrier Sense Multiple Access (CSMA)* which attempt to avoid this futile collision. In CSMA when a node wishes to transmit a packet, it first senses the channel to determine if the channel is busy. If the channel is idle, the node transmits the packet immediately. Collisions only occur if two or more nodes sense the channel idle and begin transmitting within one *propagation time*, τ, of each other. If the channel is sensed busy, the node refrains from transmitting the message and executes one of the following variants of the protocol.

1-persistent — The node waits until the channel becomes idle and transmits with probability 1.

p-persistent — The node waits until the channel becomes idle and transmits with probability p. If the channel remains idle for a propagation delay then the node again transmits its packet with probability p. If the channel becomes busy, the node repeats the protocol after waiting a randomly selected period of time.

non-persistent — The node reschedules sensing the channel to a random time in the future.

The magnitude of the propagation delay τ is critical to the performance of the protocol. If τ is small, the probability of a collision is also small since a transmitting terminal is heard by all the nodes in the network in a very short period of time. On a satellite channel however, where the propagation delay is large (on the order of 1/4 second), a terminal transmitting on the channel after hearing it idle, will not be heard until 1/4 of a second later. During this time there is a large probability that other terminals will sense the channel, hear it idle, and also transmit their packets on the channel thus resulting in a collision. The throughput of non-persistent CSMA for propagation delays typically found in land based packet radio networks is about 87% of the bandwidth of the channel.

Tobagi also studied the single-hop hidden environment and devised a protocol, *Busy-Tone-Multiple-Access* [Toba75a] in which a central station, assumed to be within line-of-sight of all terminals in the network, emits a signal indicating if the channel is busy or idle. In this way all terminals by listening to the busy tone share the same channel status information and can use this information to control their transmissions. This is actually a simplification of Tobagi's model since he considered problems associated with correctly detecting the busy tone in the presence of noise, and therefore nodes do not always have the same channel status information. It should be noted, that operating the CSMA protocol in the hidden environment without the busy tone concept, results is severely decreased throughput since nodes sensing an idle channel will transmit, and have the possibility of colliding another transmission already in progress. We

will derive equations for the maximum throughput of CSMA in the multi-hop environment in Chapter 3 of this dissertation.

Multi-hop - Silvester [Silv80a] considered a packet radio network consisting of randomly placed terminals on the plane, with finite transmission radii, using slotted-ALOHA. His derivation assumed the existence of a routing algorithm which would send, for each transmission, packets to terminals that were closest to the final destination. His results showed that to maximize the throughput of the system, terminals should transmit with a radius such that their transmissions would, on the average, be heard by approximately six other nodes. Because of this finite range, there could be several simultaneous successful transmissions in the network during any given slot, and this phenomena was named *spatial re-use* of the channel. In Chapter 2 we generalize this model to include capture and find optimal operating points for the network.

Akavia [Akav78a] studied a multi-hop model in which a continuum of terminals used slotted ALOHA to send packets to neighboring terminals. Transmitters in his model are assumed to have perfect control over the range of their transmission and receivers are assumed to be equipped with perfect capture. Let us classify the traffic into a node to help us explain some of his results. Define N to be the average number of packets in the queue of a terminal, T to be the average time packets spent in the queue, and λ to be the average arrival rate of packets to the queue. By Little's result [Litt61a] we know that $N = \lambda T$. We say that the traffic into a terminal is *bursty* if $N << 1$ and is *steady* otherwise. By using an ad-hoc formula for the delay of ALOHA networks, Akavia showed that if traffic into a terminal was bursty, terminals should transmit with enough power to reach their destinations in one hop. On the other hand, for steady traffic he found there is an optimal transmission range that minimizes the delay of the system, and in this case packets require several hops before reaching their final destinations. In another approach throughput was increased in the single-hop environment by dividing terminals into groups with different transmission powers and assuming the transmissions from low power groups do not interfere with those of higher power. It can be shown that the throughput of the ALOHA channel increases and as the limit of the number of power groups goes to infinity, it is shown the throughput of the system goes to one.

Tobagi considered the 2-hop environment in a series of papers [Toba80a, Toba80b] where he analyzed two topologies for different access schemes. In both topologies, a set of N repeaters are assumed to receive packets from N independent sets of packet radio networks each consisting of an infinite number of terminals offering traffic at a Poisson rate. Repeaters send this traffic to a central station *CS*, and only terminal-to-CS traffic is considered. The two topologies arise in the manner in which repeaters are assumed to be connected to the CS. In the *Star* topology, repeaters are assumed to have a link to the CS but are not connected to any other repeater, and in the *Fully Connected (FC)* topology all repeaters and the CS are assumed to be mutually connected. Repeaters initially are assumed to have a single buffer for storing packets.

In the first model, terminals use slotted-ALOHA to send packets to their repeaters. Upon successful reception of a packet, a repeater forwards the packet to the CS according to one of two protocols. The first protocol, Delay First Transmission (DFT), operates by having repeaters with packets to send, transmitting them with probability p in each slot. In the second protocol, Immediate First Transmission (IFT), nodes transmit newly arrived packets with probability 1 in the next slot, and collided packets are re-transmitted in the following slots with probability p. This variation, which avoids the initial packet delay, would intuitively decrease the overall packet delay from the repeater to the CS in low traffic conditions. This is verified in Tobagi's paper. Tobagi observed that although increasing the buffer size from one to two packets did improve the delay of the system, further increases in buffer size had only a minor effect. The system is said to be *channel bound* rather than *storage bound.*

In the second model, the FC topology is studied where terminals use non-persistent CSMA and the repeaters use the IFT protocol. Because of the increased throughput of CSMA over slotted-ALOHA, the throughput of the two-hop system is also increased. Since successful reception of a packet by a repeater, say *a*, implies in the FC case that no other repeater is transmitting, the channel in the next slot will be idle if *a* does not transmit its received packet. With IFT however, *a* does transmit thus taking advantage of this information, and a collision will occur only if at least one other repeater received a packet simultaneous to *a*. The probability of this is low and it is shown that the delay suffered by a packet in going from a repeater to the CS is about one packet transmission time. Once again the system is not storage bound and increasing the buffer size has little effect on the performance of the system.

Capture - In the paper by Roberts [Robe72b], power-capture is defined and the throughput for a single-hop slotted ALOHA line-of sight system is derived. It is shown that capture increases the usable bandwidth of the channel and delay equations are derived for satellite and ground packet radio systems. It is also shown that as the capture ratio varies from no-capture to perfect-capture, the throughput of the system increases.

Fratta and Sant [Frat80a] considered a packet radio network with repeaters and regular packet radio units distributed randomly on the plane. Using a power-capture model they derived lower and upper bounds on the throughput per unit area of the system as a function of the traffic offered per unit area. It is shown that the bounds for this throughput are a function of the traffic per area and the density of the repeaters. Traffic is assumed to originate at the nodes and after being transmitted to the repeaters, is sent, by some unspecified means to a central station. In a second model, a network of randomly placed nodes is studied for throughput characteristics in the single-hop line-of-sight environment. Under the assumption that each node transmits with the same probability in each slot, and that they transmit to all other nodes with equal probability, it is shown that the system has an asymptotic throughput of 1 packet per slot.

Davis and Gronemeyer [Davi80a] studied a spread spectrum system with time-capture. They studied a slotted ALOHA line-of-sight system, and developed a clever way to prevent terminals close to a given receiver from "stealing" the channel from terminals further away. This stealing would occur because their proximity to the receiver would give them an unfair advantage of being received first, and thus be more likely to be captured over more distant nodes. By causing terminals to delay transmitting for a period of time linearly dependent upon their distance to the receiver, and by adding to this a randomly selected interval of time, it is shown that all nodes will have transmission times uniformly distributed over some interval. Thus each node, regardless of its distance to the receiver, has an equal chance of being the first to be received.

1.5 Contributions of this Dissertation

In this section we outline the main contributions of this dissertation. In Chapter 2 we analyze the set of packet radio networks in which nodes are randomly distributed on the plane to determine throughput characteristics of the network as a function of transmission probability and range of transmission. This network topology corresponds to a mobile packet radio network and the shared resource we consider in this chapter is the *spatial-capacity* of the radio channel. In our model of the network, nodes use slotted-ALOHA to access the channel and have the ability to capture signals. We derive the following equation for the throughput of the network:

$$\gamma(\beta,N,p,) = \frac{45}{64}\sqrt{n\,N}\ (1-p)(1-e^{-N/2})\ p\ e^{-Np}\ Q$$

where $Q \triangleq \frac{\beta\sqrt{\beta}}{Np}\sum_{j=1}^{\infty}\frac{(4Np)^j j!}{(2j+1)!} + \frac{2}{3}\ (1-\beta\sqrt{\beta})$, n is the number of nodes in the network, N is the average number of neighboring nodes that are within the transmission range of a randomly selected node, p is the probability of transmitting in any random slot, and β is the capture parameter of the receivers of the network (recall a signal is captured if the ratio of its power to that of the other signals on the channel is greater than β^{-1} and that β=0 is non-capture and β=1 is perfect-capture). We can see that the throughput of the network increases as $\sqrt{n}$ and this factor represents the spatial-capacity of the channel. This dependency results from the fact that, because a transmission is heard by only a small subset of the nodes in the network, several transmissions can be simultaneously successful in any given slot. Typical receivers in such networks have a capture parameter of β=0.7. For this value, the point on the (N, p) plane that optimizes throughput can be found to be N=4.99 and p=.21 at which point the throughput of the network is given by $.074\sqrt{n}$. If we define the offered load, G, of the system

to be the average number of packets presented to the channel from an area equal to that of a transmission radius, then our results show that $G \cong 1$ optimizes performance. This is a generalization of the same result for the single-hop environment. In essence this model corresponds to finding optimal solutions to the interference problem using a minimal set of information.

We extend these results in the third chapter by determining an upper bound for the spatial-capacity of random networks over all protocols in a mobile environment. The goal is to free the model of the second chapter from the slotted-ALOHA assumption to determine which insights obtained from that model generalize. In particular, we find that the maximum expected fraction of successful transmissions in a connected random planar packet radio network, over all protocols, is upper bounded by:

$$f^*(N) = .9278/N$$

where, again, N is the average number of neighboring nodes that are within transmission range of a randomly selected node. This is a good rule of thumb to use in judging the performance of other protocols with these network assumptions. When we evaluate the performance of our results for optimal slotted-ALOHA against this upper bound, we find that this protocol achieves about $1/e$ of the performance given by the above equation. Recall that this is equal to the efficiency of slotted-ALOHA in a single-hop environment, and thus we have the intuitively pleasing result that the maximal capacity of slotted-ALOHA is the same in single-hop and multi-hop environments.

In the fourth chapter we seek to find the average delay messages encounter in passing through a network using a collision-free protocol. This causes us to define a new channel access protocol, called *spatial-TDMA*, for multi-hop environments in which nodes are assumed to be stationary. This corresponds to solving the interference problem with full global information. One advantage of using broadcast radio in such an environment lies in the fact that the capacity of the links in such a network can be dynamically allocated in an adaptive manner. Allocating this capacity is a resource sharing problem which we also address in chapter four. In particular we determine an approximation to the average delay that messages experience in passing through the nodes of a network using such a protocol. We determine that the delay is highly dependent upon the total time in a cyclic time frame during which specified nodes in the network are allowed to transmit, and also upon the ordering of such transmissions. This leads us to define a capacity assignment problem which attempts to minimize the average message delay for all messages in the network over all possible orderings and lengths of transmission periods. Although this problem is mathematically intractable, we do find such optimal operating points for networks in which the ordering of the intervals has been randomly assigned. This is a upper bound on the delay that would be obtained for the best possible ordering.

In the fifth chapter we seek to determine optimal use of the information that nodes obtain when they sense the channel in their local environments. This causes us to define a new class of protocols called *rude-CSMA* which use this information. We formulate the protocol, derive equations for its performance, and optimize it over different classes of networks. This investigation corresponds to solving the interference problem using only local information. In these protocols we assume that nodes, after sensing the channel, can estimate the number of transmitting nodes that are within their hearing distance. Intuitively, if this number is large, there is high probability, if the sensing node transmits a packet, that its message would undergo a collision. If however, there was only one transmitter, say, in its local neighborhood, then there exists the possibility that the sensing node would be sending a packet to a radio unit that was outside the transmitting node's range, and it would be successfully received. Thus transmitting with a non-zero rate might increase the number of simultaneous successful transmissions in the network. This motivates the protocol and explains why we have called it "rude-CSMA", since nodes sometimes transmit even when they sense the channel busy. If we define the state $S = (s_1(S), s_2(S), \cdots, s_n(S))$ of the system to be a binary vector in which $s_i(S)=1$ if node i is transmitting a packet and equal to 0 otherwise, then the rate at which node i presents packets to the channel (assuming that $s_i(S)=0$) in state S is given by:

$$r_0^i(S) = \gamma_0 \, x^{N_0^i(S)} \, y^{N_1^i(S)}$$

where γ_0 is the arrival rate of messages to each node of the network, $N_0^i(S)$ is the number of neighboring nodes of node i that are not transmitting, $N_1^i(S)$ is the number of transmitting neighboring nodes, and x and y are parameters of the protocol. On the other hand if node i is a transmitter, $s_i(S)=1$, then the rate at which its transmission stops is given by:

$$r_1^i(S) = \mu$$

which corresponds to exponential packet lengths. One sees that the protocols defined by these rates have ALOHA ($x=1$, $y=1$) and CSMA ($x=1$, $y=0$) as special cases and thus should have an optimal performance as good as either of them. It can be shown that the steady state probability of state S is given by:

$$\Pi(S) = C \, \rho^{M(S)} \, x^{-B_0(S)} \, y^{B_1(S)}$$

where C is a normalization constant, $M(S)$ is the number of transmitters, $B_0(S)$ is the number of neighboring nodes that are not transmitters and $B_1(S)$ are the number of neighboring transmitters for state S. Using these equations we analyze specific networks to determine the parameter values (x, y) that achieve maximal throughput. In particular, we find that over all random topologies that we studied, $y=0$ had maximal performance and thus nodes should engage in a type of CSMA protocol (never transmit when the channel is sensed busy). For lattice type networks we found that, for certain input values, throughput was increased if nodes after sensing a busy channel, transmit with a non-zero rate (i.e. $y \neq 0$). It thus turns out that, in real networks (which would have topologies more random-like), the binary information obtained from sensing the channel is sufficient to determine if one should transmit or not.

We have adopted what we hope will be a clear writing style in this dissertation. Chapters have been made as self-sufficient as possible to allow the reader to selectively read portions they find most interesting without the necessity of reading all the previous material. This of course, implies that there is a certain amount of redundancy and repeated definitions in the text, but it is hoped that these repetitions will serve to clarify the material rather than to bore the reader. A list of symbols and notations which are used throughout the dissertation is included in the beginning pages and all notations used in a chapter are defined in the introductory sections of the chapter.

1.6 Applications of Packet Radio Networks

The main advantages of packet radio networks over conventional networks is that they are not dependent on fixed topologies, are easy to establish, and can operate unattended. These characteristics allow terminals to be mobile and to be attached to a diversity of computing and sensor devices. Many interesting new applications of this technology can be found in [Lick78a]. In one such application, sensors are carried by elderly persons which monitor their physical orientation while walking. If the sensor detects a horizontal position (implying they fell down) it notifies the client's home computer which telephones a hospital with a preprogrammed message and the location (as determined by the local radio network) of the client. When medics arrive, they can solicit the help of a physician located, miles away, by attaching sensors to the client's body that transmit data the doctor can use to control and monitor the treatment given. Any number of other applications using unattended sensors that communicate with computers to control and monitor specific activities such as alarm systems for security, fire, and gas or water leakage, are natural applications of such networks.

The fact that it is not necessary to use wires to connect devices together allows easier implementation of communication systems. One can imagine a hierarchical network in which the lowest level is a local community radio network that transmits packets to a central subnetwork station. These packets would then be relayed along intra-city high bandwidth lines to a main switching station that connects with a long haul inter-state network having links to international satellite facilities. Since radio is used on the local level, there is no necessity for erecting poles and stringing cable directly to subscriber's equipment. Besides beautifying the landscape of our cities, this also has the advantage that providing service to a new subscriber is as easy as outfitting them with proper radio equipment. Both voice and data would pass digitally over this network thus supporting a mobile telephone system and providing access to a wealth of data and computing resources resident on the net.

Since packet radios will be inexpensive and about the size of a pocket calculator [Kahn78a], we would imagine their use will be widespread and thus revolutionize the manner in which the society negotiates its information transactions. No longer will we have to physically wait in lines to negotiate our banking business, for instance, when such transactions could be performed from the convenience of our homes (or even our cars!) and some information in the present form of newspapers, magazines, books, stereo records, photographs and video could be sent digitally over the network, stored on disk, and recalled at our convenience.

In emergency situations, say as in an earthquake disaster, the flexible topology of packet radio networks is necessary to establish communications, since in such situations land based lines will likely be destroyed. The immediate adaptation of such networks to changing topologies is critical to the performance of such an operation in which new rescue teams continually arrive with more packet radio units; and the fact that it is easy to change the responsibilities of a radio unit from a regular packet radio unit to an unattended repeater, allows a multi-hop network to be configured rapidly thus increasing the useful range of these networks.

CHAPTER 2
OPTIMAL THROUGHPUT FOR SLOTTED-ALOHA IN A MOBILE NETWORK

In this chapter we derive the throughput equation for a mobile packet radio network in which nodes use slotted-ALOHA to access the channel. We then optimize this equation over certain parameter values to find the optimal throughput for these networks. This corresponds to solving the interference problem using a minimal amount of information.

2.1 Introduction

In this chapter we analyze a basic trade-off in multi-hop radio networks which concerns the throughput of the network. Recall that throughput is defined to be the number of packets reaching their final destinations per unit time. To illustrate the problem discussed in this chapter we will consider a network under two different cases. First suppose that all nodes of the network transmit with a very small power. This implies that packets must be relayed over several hops before reaching their final destination and thus tends to decrease the throughput of the network. We cannot however conclude that increasing the range of the transmitters of the network will increase the throughput of the network. For suppose nodes do have an increased range of transmissions. Then messages must be relayed fewer times, but at each hop there is a higher probability that the transmission is not successfully received due to the increased amount of interference caused by each transmission. This then increases the number of collisions in the network and tends to decrease the throughput.

The tradeoff can thus be summarized as: increasing the power at which a node transmits decreases the number of hops needed to reach the final destination but causes more interference at each hop, while decreasing the power has the corresponding opposite effects. Finding the operating point that maximizes the throughput for a random network is the problem that is discussed in this chapter and corresponds to solving the interference problem in which nodes use a minimal set of information (namely the density of terminals on the plane).

2.2 The Models

Throughout this chapter we will make the following assumptions about the network:

a. *Topology* - We assume that packet radios are distributed according to a Poisson point process on the plane with a mean density of λ packet radio units (also called terminals) per unit area. We are interested in finding the throughput for an area containing a large number, n, of packet radios and will ignore edge effects. This topology represents an instantaneous snapshot of a large mobile packet radio network.

b. *Stations* - We assume that each packet radio transmits with fixed power, and that all n stations (i.e. terminals) in the network transmit with the same power on the same frequency band. Receivers are assumed to be able to receive a signal from another station if that station is within a radius R of the transmitter, and under certain circumstances, can successfully capture one of several simultaneous transmissions within its hearing range. Let us refine the definition of capture given in the introduction by considering a receiver a which is within range of two transmitters t_1 and t_2 and assuming that t_1 has a packet destined for a. Let P_1 and P_2 be the powers of the signals received by a, and r_1 and r_2 be the distances between a and the two transmitters. Whenever both t_1 and t_2 transmit their packets simultaneously, their signals interfere with each other. In the absence of capture, station t_1 will not be received correctly. With capture however, station a can successfully receive t_1's transmission if $P_1/P_2 > \beta^{-1}$ $(0 \leqslant \beta \leqslant 1)$, where β is called the *capture-ratio*. Assuming omni-directional antennas on the plane and equal transmitting power for all stations, this ratio of powers can be converted into a ratio of distances since the power of a received signal decreases as the inverse square of the distance. Thus, using this distance measure, a will capture t_1 if $r_2/r_1 > \beta^{-1/2}$. From Figure 2.1 we see that t_1 will be successful if t_2 lies outside the circle of radius $r\beta^{-1/2}$, called the *capture radius*. Observe that $\beta = 0$ implies that simultaneous transmission will always cause a collision *(non-capture)*, and $\beta = 1$ implies t_1 will be received if it is simply closer to a than t_2 *(perfect-capture)*. Well designed FM receivers have a capture ratio approximately equal to 0.7 [Robe72b]. Although the appearance of the exponent $(-1/2)$ of β in the above equation appears awkward, this selection simplifies later equations. The last assumption we make about the stations is that each radio is always busy, and thus we study the heavy traffic case.

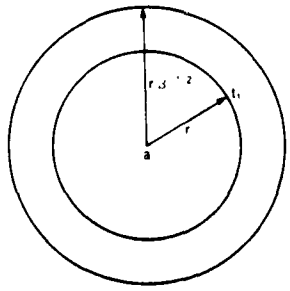


FIGURE 2.1.
Definition of Capture.

c. *Channel Access Method* - We assume that the time axis is slotted and that the probability a station will transmit a packet on the channel in any given slot is p.

d. *Traffic Matrix* - Since nodes are Poissonly distributed on the plane with mean density λ, and since any station can send and receive packets within a radius R, every station has on the average $N = \lambda \pi R^2$ neighbors (terminals within its hearing and transmitting range). We assume the global traffic matrix for all the n nodes in the network is uniform, and thus the probability of sending to any particular node in the network is $1/n$.

e. *Routing* - We choose to study the case where packets destined towards a particular node F in the network are routed with equal probability towards one immediate neighboring node that lies in the general direction of F. For example in Figure 2.3 there are k terminals within transmitting range of t that are closer to F than to t. Transmitter t will pick one terminal from these k neighbors with probability $1/k$. Suppose t transmits to node i. We call the difference of the distances between t and F, and i and F, the *forward progress* of the transmitted message.

Let us justify the random routing assumption by comparing it to an optimal routing model. In such a routing policy packets are assumed to be relayed to a neighboring terminal that is closest (furthest along the path) to that packet's final destination. Such an optimal routing policy is not realizable for a mobile packet radio network since it requires the exact location of all terminals (which are assumed to be moving) as well as that of the final destination, and would provide an upper bound for network performance. This upper bound can be easily calculated in networks without capture because the probability of being successfully received is independent of the distance between the transmitter and receiver. In the capture environment however, nodes closer to the transmitter have a greater probability of receiving a transmitted signal than those further away. This non-uniformity makes the calculation of the distance covered in one transmission for an optimal routing policy difficult.

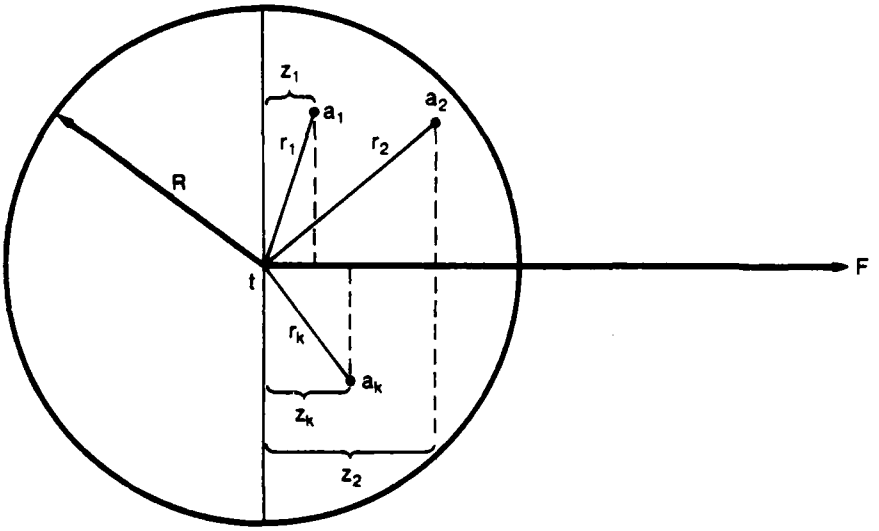


FIGURE 2.2.
The Routing Assumption.

To be specific, suppose a transmitter t has a message to send to a particular final node F (Figure 2.2). Suppose t has k neighbors lying in the half circle of his transmission radius, towards F, and that their distances from t are $(r_1, r_2, \ldots, r_k)$. Let $(z_1, z_2, \ldots, z_k)$ be the vector of projected distances towards F, hence if t transmits to node i at r_i, the progress towards F will be z_i. Let $P(r_i)$ be the probability that node i successfully receives t's transmission. A locally optimal routing algorithm for this system is defined as one that sends all packets towards F to the node j that has the maximum value of $z_i P(r_i)$. To determine this value, one must calculate the joint probability for $(r_1, r_2, \ldots, r_k)$ and $(z_1, z_2, \ldots, z_k)$ for all k, to determine the density for

the maximum projected distance. This would then be unconditioned on k to determine the density for the maximal forward progress. In the non-capture environment $P(r_i) = P(r_j)$ for all i and j, and thus maximizing $z_i P(r_i)$ implies picking the maximum z_i. Over all sets of k nodes, the probability that the maximum projected distance is equal to a certain value, say z, is seen to be the probability that there are no terminals in the half circle from t that are closer to F (the shaded region A in Figure 2.3).

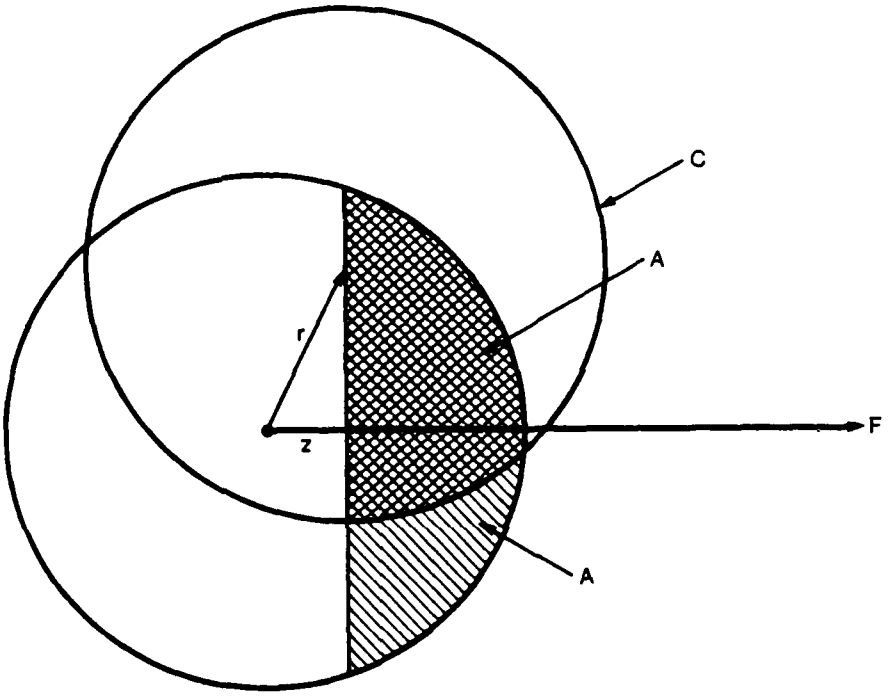


FIGURE 2.3.
Probability of Successful Reception.

Since terminals are Poissonly distributed on the plane, this probability equals $e^{-\lambda A}$. In the capture environment however, $P(r_i) = P(r_j)$ only if $r_i = r_j$, and the above calculation is no longer valid. The optimal routing algorithm now will no longer always pick the node with the maximum z_i, because the product $z_i P(r_i)$ may not be maximal. To avoid these computational complexities and in endeavoring to create a practical bound for packet radio networks, we have chosen to assume a random routing policy (assumption e. above). Random policies have been proposed for networks of this kind [Klei64a, Liu80a] and our calculations will be a lower bound on the performance for algorithms that always send packets in the direction of their final destination.

With this random routing assumption we will analyze two models that differ in the way capture is defined in relation to the maximum transmission distance R. Continuing with the scenario of the previous section, it is certainly true that in the presence of multiple simultaneous transmissions receiver a can only successfully receive packets from its closest transmitting neighbor. Thus we can say that a receives from t_1 if t_1 is a's nearest transmitter, and there are no other transmitters within the capture radius. Let us therefore suppose in our scenario that t_1 is a's closest transmitter and is located a distance r away. Our two models differ in the way they define the capture radius in relation to the maximum transmitting distance R. Suppose $r\beta^{-1/2} > R$, or in words, that the capture radius is greater than a's maximum hearing distance. Certainly a transmitter located further than R, say at r' such that $R < r' < r\beta^{-1/2}$ will have no effect on a's reception since his signal will be too weak to be received. Thus in Model 1 we define the capture radius to be equal to the minimum of $r\beta^{-1/2}$ and R. The area that must contain no other transmitters for a to successfully receive t_1's transmission, the *clean area*, for Model 1 is the annulus of inner radius equal to r and outer radius equal to the minimum of $r\beta^{-1/2}$ and R. This definition for the capture radius for Model 1 however gives weak signals coming from a transmitter located at a distance slightly less than R, say at $R-\epsilon$, a greater probability of being successfully received than a transmitter with a smaller value of r, since the clean area, the annulus of inner radius $R-\epsilon$ and width ϵ, is infinitesimal. In practice however, if there was a transmitter at a distance slightly greater than R, say at $R+\epsilon$, it would disrupt reception since the ratio of the powers of the two transmitters would be close to 1 even though the second transmitter's signal was very weak. Model 2 attempts to account for this discrepancy by defining the clean area to be $r\beta^{-1/2}$ regardless of the relationship between $r\beta^{-1/2}$ and R. We observe since the clean area for perfect capture ($\beta = 1$) is identical in both models, we would expect our equations to the same for this case.

We must comment that both models make two simplifying assumptions about the capture phenomenon. In actual practice, one particular transmitter, say t_1, will be captured by a certain receiver if the ratio of its received power, to the sum of the received powers of *all* other signals simultaneously heard by the receiver, is greater than the specified capture ratio. Letting P_{t_i} be the receiver power for the i'th transmitter, k be the number of transmitters the receiver hears, and assuming that the powers are sorted into decreasing order ($P_{t_1} > P_{t_2} > \cdots > P_{t_k}$), we have that t_1 is captured if $P_{t_1} / \sum_{i=2}^{k} P_{t_i} > \beta^{-1}$. In our models we approximate the sum of all the powers of terminals t_2 through t_k by the power of the next strongest signal t_2. This assumption however, is not crucial as we will see later, since the optimal choice of system parameters tends to separate transmitting stations so that on the average if a receiver hears more than one terminal transmitting, with high probability it hears exactly two transmitters and therefore the above sum and P_{t_2} are identical. The second simplifying assumption we make is that capture is a deterministic phenomenon such that if the ratio of the received powers is greater than β then the signal is captured with probability 1. In

actuality however, capture is probabilistic and has a density that is a function of the ratio of the received powers and of the capture parameter. The results of our deterministic model can be applied to this more realistic model however, without too much error, by using a value of β so that if $P_{t_1} / \sum_{i=2}^{k} P_{t_i} > \beta^{-1}$ then the actual probability of being captured is greater than some specified confidence probability (say 0.95).

2.3 Analysis of Model 1

2.3.1 Expected Number of Successful Receptions

We first calculate the probability of successful reception for a randomly selected terminal in the network. Let us assume that terminal a captures the transmission of its closest transmitting neighbor t. Conditioned on this, a will successfully receive t's packet if the packet was addressed to a and if a did not transmit in the current slot. This occurs with probability

$$P[E_s | \textit{no interference}] = \frac{(1-p)\,(1-e^{-V/2})}{N}$$

where E_s is the event of a successful reception.
We can see this by first defining the following events

E_t - The event that t sends to a.

E_d - The event that t sends in the direction of a.

$N(i)$ - The probability that there are i other terminals besides a in the half circle of radius R from t.

In Figure 2.2 for example their are k terminals in the half circle from t towards F, and transmitter t is sending in the direction of all the labeled terminals in the figure.

We know that

$$P[E_t] = P[E_t|E_d]\,P[E_d] + P[E_t|E_d^c]\,P[E_d^c]$$

but $P[E_t|E_d^c] = 0$ since we do not allow packets to go away from their destination. Since t's destination is uniformly distributed over the plane, a randomly chosen terminal a within a radius R of t is in the direction of F, is equal to the probability that a lies in the half circle of radius R directed from t to F. Since a is equally likely to be in either half we have $P[E_d] = 1/2$. Calculating the remaining term, $P[E_t|E_d]$, we have

$$P[E_t|E_d] = \sum_{i=0}^{\infty} P[E_t|E_d, N(i)]\,P[N(i)|E_d]$$

Each of these terms is known, for $P[N(i)|E_d] = P[N(i)]$ is Poisson with parameter $\lambda\pi R^2/2$ and $P[E_t|E_d, N(i)] = 1/(i+1)$, since if there are i other terminals besides a, making a total of $i+1$ terminals, t will select one of them with equal probability. Recalling that the average number of neighbors is $N = \lambda\pi R^2$, we have:

$$P[E_t|E_d] = \sum_{i=0}^{\infty} \frac{1}{i+1}\,\frac{e^{-N/2}\,(N/2)^i}{i!} = \frac{2}{N}\,(1-e^{-N/2})$$

Combining with the previous calculations we obtain:

$$P[E_t] = \frac{1}{2}\,P[E_t|E_d] = \frac{(1-e^{-N/2})}{N}$$

Knowing that the packet was addressed to a in the absence of interference we know that a will successfully receive t's packet if a does not transmit, thus giving the $(1-p)$ term and establishing the above expression.

We must now calculate the probability that there is no interfering traffic. We do this by first conditioning on the distance between a and t to be r $(r \leqslant R)$, and then analyzing two cases:

Case 1. $\beta^{-1/2}r < R$ (Figure 2.4a)

In this case a will receive the packet if there are no other transmitters in $(r, r\beta^{-1/2})$, the *clean area*. This area is equal to $\pi(\beta^{-1/2}r)^2 - \pi r^2 = \pi r^2(1/\beta - 1)$ which contains no transmitters with probability $e^{-\lambda p\pi r^2(1/\beta-1)}$.

Case 2. $\beta^{-1/2}r > R$ (Figure 2.4b)

The area that must now be clean is seen to be $(\pi R^2 - \pi r^2)$ which occurs with probability $e^{-\lambda p\pi(R^2-r^2)}$.

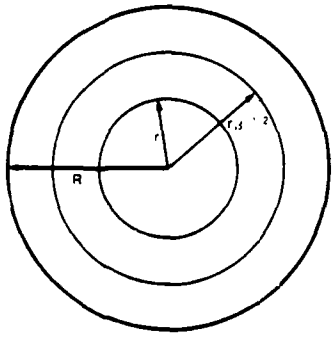


FIGURE 2.4a.
Capture: Case 1.

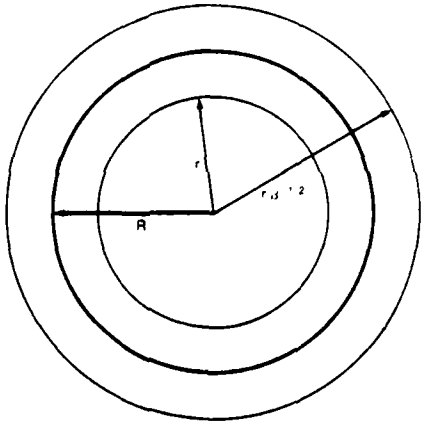


FIGURE 2.4b.
Capture: Case 2.

The density for the distance between a randomly selected terminal and its closest transmitting neighbor, can be easily calculated. Letting X be the random variable for this distance, and knowing that busy terminals are Poissonly distributed on the plane with parameter λp we have

$$P[X \leqslant r] = 1 - P[\textit{no busy terminal in } (0,r)] = 1 - e^{-\lambda \pi p r^2}$$

Hence letting $f(r)$ be the density for X

$$f(r) = \frac{d\,P[X \leq r]}{dr} = 2\pi\lambda p\, r\, e^{-\lambda p \pi r^2}$$

Using this in the above, after unconditioning we obtain:

$$P[E_s] = \frac{(1-p)(1-e^{-N/2})}{N}\left[\int_0^{R\beta^{1/2}} e^{-\lambda p\pi r^2(1/\beta-1)}\, 2\pi\lambda pre^{-\lambda p\pi r^2}\, dr \right.$$

$$\left. + \int_{R\beta^{1/2}}^{R} e^{-\lambda p\pi(R^2-r^2)}\, 2\pi\lambda pre^{-\lambda p\pi r^2}\, dr\right]$$

Performing the integration we have:

$$P[E_s] = \frac{(1-p)(1-e^{-N/2})}{N}\left[\beta(1-e^{-Np}) + (1-\beta)Npe^{-Np}\right]$$

For future equations let $Y \triangleq \beta(1-e^{-Np}) + (1-\beta)Npe^{-Np}$.

We will discuss two special cases to see that they are intuitively plausible.

Case 1. $\beta = 0$ the non-capture case for which

$$P[E_s] = (1-p)(1-e^{-N/2})\, p\, e^{-Np}$$

Intuitive explanation - For a to receive, it must not transmit and this occurs with probability $(1-p)$. Receiver a also cannot be isolated from other nodes in the network. In particular one half circle of radius R must contain at least one other terminal and this occurs with probability $(1-e^{-N/2})$. Out of the neighbors in this region only 1 can transmit (occurring with probability Npe^{-Np}), and on the average that transmitter is surrounded by N neighbors and thus transmits to receiver a with probability $1/N$. Combining all the above probabilities we obtain the same expression as above.

Case 2. $\beta = 1$ the perfect capture case where

$$P[E_s] = \frac{(1-p)(1-e^{-N/2})}{N}(1-e^{-Np})$$

Intuitive explanation - Again a must be silent which occurs with probability $(1-p)$, and must not be isolated from nodes in one half circle of radius R which occurs with probability $(1-e^{-N/2})$. There must be at least one transmitting station in its neighborhood (occurring with probability $1-e^{-Np}$) and since only a's nearest neighbor can be successfully received by a, the probability it transmits a packet to a on the average is $1/N$.

To calculate the expected number of successes in the network (denoted by E) we merely have to multiply the previous probability by the number of nodes n in the network to obtain:

$$E = n\,P[E_s] = \frac{n}{N}\,(1-p)(1-e^{-N/2})\;Y \tag{2.1}$$

We can check this equation against the well know single-hop slotted-ALOHA results for the infinite population model with Poisson traffic statistics. We do this by setting, in equation (2.1) $n = N$ and $\beta = 0$, and then performing two limit operations. By letting $p \to 0$ as $n \to \infty$ in such a way as to preserve the product $G = np$ to be a constant, we obtain Poisson traffic characteristics with parameter G. We then obtain

$$\lim_{\substack{n=N\to\infty \\ p\to 0 \\ G=np=constant}} \frac{n}{N}\,(1-p)\,(1-e^{-N/2})\,N\;p\;e^{-Np} = G\,e^{-G}$$

This reaches its maximum at $G=1$ giving the familiar maximum throughput for the slotted-ALOHA channel [Abra77a] of $1/e$.

It can be seen that the function E is increasing in β, the capture-parameter, by writing E as a linear function of β:

$$E = H(\beta) = K\,(\beta\,(1-e^{-Np}\,(1-Np)) + N\;p\;e^{-Np})$$

where $K = \frac{n}{N}(1-p)(1-e^{-N/2})$. This describes a straight line with slope $m = 1-e^{-Np}\,(1-Np)$. But this slope is always positive since $m < 0$ implies $e^{Np} < 1-Np$ which is seen to be false by expanding e^{Np} in a Maclaurin series. We conclude that increasing the receiver's ability to capture signals increases the expected number of successes in the network.

Recall that Y, as defined above, is the probability that there is no interfering traffic to a's successful reception. We see that $Y = \beta(1-e^{-Np}) + (1-\beta)Npe^{-Np}$ can be viewed as a convex combination of β since $0 \leqslant \beta \leqslant 1$. We cannot resist interpreting β then as a probability (mathematicians beware !) to see if this interpretation provides some intuitive insight into our problem. If we interpret β as the probability of capturing signals, then for the non-capture case, $\beta=0$, we have that the probability that a has no interference is given by:

$$Y_{\beta=0} = Npe^{-Np}$$

This however, is the probability that there is exactly one transmitter within a's local environment which is a necessary condition for non-interference with non-capture. Likewise when $\beta=1$, the perfect-capture case, we have:

$$Y_{\beta=1} = 1-e^{-Np}$$

which is the probability that there exists at least one transmitter within a's local environment. This of course, with perfect-capture, is a necessary condition for successful reception since non-interference is certain (a always captures its nearest transmitter). We thus have the intuitively pleasing result that we can view β as being the probability of capturing signals.

We now seek to determine the maximum number of expected successes in the network. Certainly this number must be less than $n/2$ since every successful receiver is associated with exactly one successful transmitter. It is easy to verify this analytically. Numerically calculating the maximum of E for various values of β demonstrates that the maximum expected number of terminals in the network that could engage in successful communication at any given slot is about 21% for perfect capture and about 14% for the non-capture environment. The exact values are reported in Table 2.1. We note here that the values of N and p that maximize the probability of success, do not also maximize the throughput of the network (see Table 2.2). This is a result of the dependency of the probability of successful transmission and the maximum transmission range R as seen in the equation $N = \lambda \pi R^2$, and will be discussed in greater length in the next section. Observe that the results of Table 2.1 would only be applicable to networks in which all packets went exactly one-hop to reach their final destination.

β	N	p	$P[E_s]$
0.0	1.9880	.29377	.07280
0.1	2.0594	.29974	.07557
0.2	2.1365	.30594	.07846
0.3	2.2195	.31239	.08154
0.4	2.3085	.31963	.08484
0.5	2.3036	.32585	.08835
0.6	2.5044	.33276	.09210
0.7	2.6102	.33970	.09609
0.8	2.7201	.34659	.10030
0.9	2.8326	.35331	.10477
1.0	2.9462	.35977	.10946

TABLE 2.1

2.3.2 Expected Number of Successful Transmissions

Since every successful packet in the network has exactly one receiver-transmitter pair associated with it, the number of successful receivers in the network is equal to the number of successful transmitters. Thus our calculation in this section should yield (2.1) of the last section. The approach in this section however, allows us to calculate the density for the distance covered in a successful transmission, and serves to support our assumption of random routing. In analyzing an optimal routing algorithm, the spatial information used in determining the next terminal in the route must also be used in calculating the probability of successfully transmitting a packet to that terminal. Such calculations in [Klei78a] where the authors assumed an optimal routing policy, were performed separately and produced an inconsistency in their results. Since a determination of the maximal transmitted distance, say z, implies in their calculation that there are no terminals in area A of Figure 2.3, the calculation of the probability of success is the probability there are no transmitters within a radius of R of the receiver. This in Figure 2.3 is equal to $e^{-p(N-A')}$ where A' is the intersection of circle C and area A. In [Klei78a] knowledge that A' contains no transmitters was not used in their calculation of the probability of successful reception. With this inconsistency the expected number of successful receivers in the network is not equal to the expected number of successful transmitters. The two ways we calculate the expected number of successes in the network do however agree and thus assure us that our assumption of random routing is consistent with both derivations.

To calculate the expected number of successful transmissions in the network over one slot, we first calculate the probability that a randomly chosen station will successfully transmit a packet to a neighboring station, and then multiply this probability by the number of stations in the net, n. Let us define:

X - a random variable denoting the distance between the transmitter and its intended receiver.

E_t - The event that the randomly chosen terminal transmits a packet in the given slot.

E_i - The event that there exists at least 1 other station in the half circle of radius R in the direction of the intended final destination.

E_s - The event of successful reception of the transmission by its intended neighboring terminal.

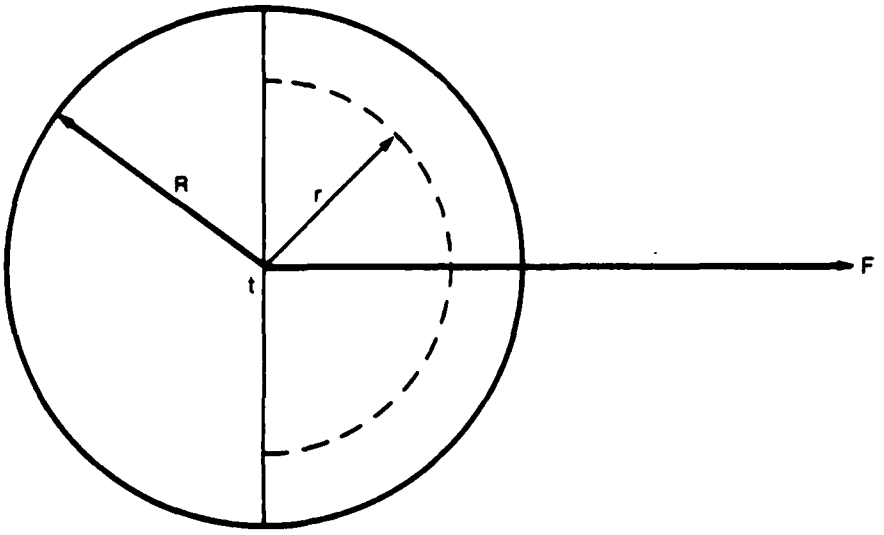


FIGURE 2.5.
Probability of Successful Transmission.

Since we assume that routing is uniform over all the nodes in the right-hand region of Figure 2.5 we have:

$$P[X \leqslant r | E_t, E_i] = \frac{\pi\, r^2/2}{\pi\, R^2/2} = \frac{r^2}{R^2}$$

Differentiating we obtain:

$$P[r \leqslant X \leqslant r + dr | E_t, E_i] = \frac{2\,r}{R^2}\, dr$$

Our definition of the capture phenomena for Model 1 implies:

$$P[r \leqslant X \leqslant r + dr, E_s | E_t, E_i] = \frac{2\,r}{R^2}\,(1-p)\; e^{-\lambda\pi p\left[\min(r\beta^{-1/2}, R)\right]^2}\, dr \qquad (2.2)$$

The probability of successful transmission, $P[E_s]$, is found after unconditioning and integrating:

$$P[E_s] = p\,(1-p)\,(1-e^{-N/2}) \int_0^R \frac{2\,r}{R^2}\, e^{-\lambda\pi p\left[\min(r\beta^{-1/2}, R)\right]^2}\, dr$$

Breaking this integral up into regions $[0, R\beta^{1/2}]$ and $(R\beta^{1/2}, R]$ we obtain:

$$P[E_s] = \frac{(1-p)\,(1-e^{-N/2})}{N}\, Y \qquad (2.3)$$

Thus the expected number of successful transmissions is found to be

$$E = \left(\frac{n}{N}\right)(1-p)(1-e^{-N/2})\ Y$$

which is the same as (2.1).

2.3.3 Expected Forward Progress

We are now in a position to derive the density for the distance between transmitter and receiver for a successful transmission. Referring back to equation (2.2) and using (2.3) we have,

$$P[r \leq X \leq r+dr \mid E_s] = \frac{P[r \leq X \leq r+dr, E_s]}{P[E_s]}$$

$$= \frac{p(1-p)(1-e^{-N/2})}{P[E_s]} \quad \frac{2r}{R^2}\, e^{-\lambda\pi p\left[\min(r\beta^{-1/2},R)\right]^2}\, dr \qquad (2.4)$$

or defining $g(r)$ to be the density for r and $\frac{1}{s} = \frac{Np}{Y}$ we may rewrite equation (2.4) in regions to obtain

$$g(r) = \begin{cases} \frac{1}{s}\frac{2r}{R^2}\, e^{-\lambda p\pi r^2/\beta} & 0 \leq r \leq R\beta^{1/2} \\ \frac{1}{s}\frac{2r}{R^2}\, e^{-\lambda p\pi R^2} & R\beta^{1/2} \leq r \leq R \end{cases} \qquad (2.5)$$

It can be easily verified that this integrates to 1 and thus is a proper density.

Suppose as shown in Figure 2.6, transmitter t is sending a packet to final destination F through intermediate node a. We wish to calculate the progress of the packet towards its final destination. To simplify the calculation we assume that forward progress will be the same for any node on the line perpendicular to the direction of the destination, line L in the figure. This assumption is reasonable if the distance D is much greater than R. Because terminals are randomly distributed on the plane, for a given distance r and a given destination F, the angle θ will be uniformly distributed over $(-\pi/2, \pi/2)$. Define Z to be the random variable denoting the forward distance. We see that for a given r the probability that Z is less than some value z is the same as the probability of $|\theta|$ being larger than $\cos^{-1}(z/r)$ or letting $F(z)$ be the

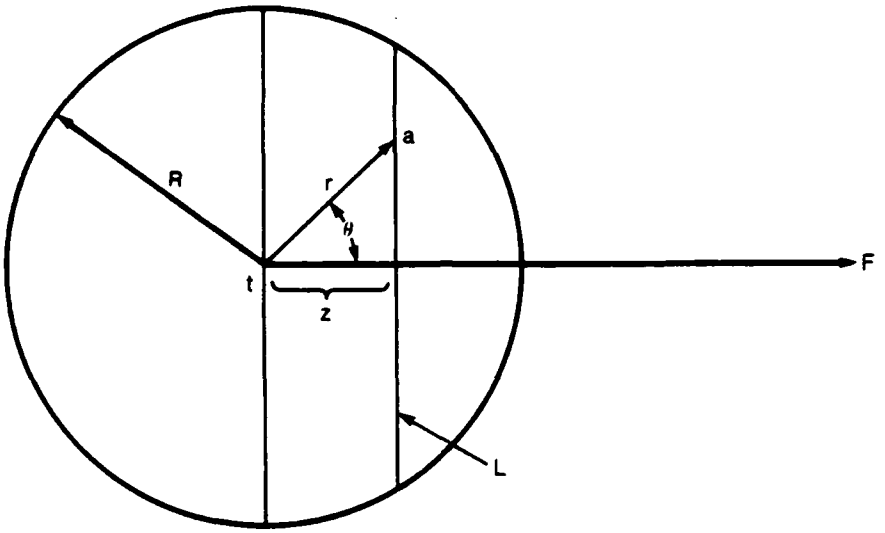


FIGURE 2.6.
Expected Forward Progress.

distribution of Z we have

$$F(z|r) = \begin{cases} 1 & r < z \\ 1 - \dfrac{2\cos^{-1}(z/r)}{\pi} & 0 \leqslant z \leqslant r \leqslant R \end{cases}$$

Differentiating this with respect to z we derive the conditional density

$$f(z|r) = \begin{cases} 0 & r < z \\ \dfrac{2}{\pi\, r\sqrt{1-z^2/r^2}} & 0 \leqslant z \leqslant r \leqslant R \end{cases}$$

We can now calculate the expected progress given r

$$E[Z|r] = \int_0^r z\, f(z|r)\, dz = \frac{2}{\pi}\int_0^r \frac{z}{\sqrt{r^2 - z^2}}\, dz = \frac{2\,r}{\pi}$$

We can uncondition this by using the density of (2.5) to obtain

$$E[Z] = \frac{2}{\pi}\int_0^R r\, g(r) = \frac{2}{s\,R^2\,\pi}\left[\int_0^{R\beta^{1/2}} 2\,r^2\, e^{-\lambda\rho\pi r^2/\beta}\, dr + e^{-\lambda\rho\pi R^2}\int_{R\beta^{1/2}}^R 2\,r^2 dr\right] \qquad (2.6)$$

The second integral within the brackets is equal to $\frac{2}{3} R^3 (1 - \beta\sqrt{\beta})\ e^{-\lambda p \pi R^2}$ and letting the first integral be denoted by *INT*, we can write equation (2.6) as

$$\bar{z} = \frac{4\ INT}{s\ R^2\ \pi} + \frac{4\ R}{3\ s\ \pi}\ e^{-Np}\ (1 - \beta\sqrt{\beta}) \tag{2.7}$$

In Appendix A it is shown that

$$\int_0^x t^2 e^{-k\ t^2} dt = x\ \frac{e^{-kx^2}}{2\ k} \sum_{j=1}^{\infty} \frac{(4\ k\ x^2)^j\ j!}{(2\ j\ +\ 1)!}$$

and thus using this for *INT* in (2.7) and reducing we finally obtain:

$$\bar{z} = \frac{2}{s\ \pi}\ e^{-Np}\ R \left[\frac{\beta\sqrt{\beta}}{Np} \sum_{j=1}^{\infty} \frac{(4Np)^j j!}{(2j+1)!} + \frac{2}{3}\ (1 - \beta\sqrt{\beta}) \right] \tag{2.8}$$

For future equations, let $Q \triangleq \frac{\beta\sqrt{\beta}}{Np} \sum_{j=1}^{\infty} \frac{(4Np)^j j!}{(2j+1)!} + \frac{2}{3}\ (1 - \beta\sqrt{\beta})$.

2.3.4 Expected Throughput

We can calculate the expected throughput for the network for each slot. For any randomly selected terminal, the expected path length between it and another randomly selected terminal is given in [Kend63a] as $d = \frac{128}{45\pi} \left[\frac{n}{\lambda\pi} \right]^{1/2}$. Since $\bar{z}$, as calculated in the previous section is known, the number of hops h a randomly selected packet will take is given by $h = d/\bar{z}$. Therefore the average number of messages delivered to their final destinations per slot, the throughput, is given by $\gamma = nP[E_s]/h = nP[E_s]\ \bar{z}/d$. Using the previously derived equations for the quantities above we obtain the final result:

$$\gamma(\beta, N, p,) = \frac{45}{64} \sqrt{n\ N}\ (1 - p)(1 - e^{-N/2})\ p\ e^{-Np}\ Q$$

The increase in the throughput γ with the square root of the number of terminals in the network is a result of the spatial reuse of the channel. Observe that the equation obeys our intuition for $p = 0$ or $p = 1$ when the throughput is zero, and that the $(1 - e^{-N/2})$ term is the probability that the network is connected over one hop. Once again we can show that the throughput of the system is an increasing function of β since if the function is increasing in $\alpha = \beta\sqrt{\beta}$ then it is also increasing in β. Thus γ is linear in α with slope $m = \sum_{j=1}^{\infty} \frac{4^j\ (Np)^{j-1}\ j!}{(2j+1)!} - \frac{2}{3}$ which

is clearly positive since the minimum of m is 2/3. Thus as we have seen previously, increasing the capture-parameter can only increase performance.

2.3.5 Discussion of Results

In all the following graphs and Tables we use normalized throughput $\gamma'(N,p,\beta) = \gamma(N,p,\beta)/\sqrt{n}$, hence eliminating the dependency of the size of the network from our equations. We must note that the square root of n dependency on the throughput of the network is an important factor that lets us achieve, by reducing the strength of transmitted signals, throughputs greater than that obtainable by running the network as a one-hop ALOHA network. For example, in Table 2.2 for $\beta = 0.7$, we see that $\gamma' = 0.0749282$. To determine the number of terminals needed in the network to achieve throughput greater than $1/e$ we set $\sqrt{n}\,\gamma' > 1/e$ which implies $n \approx 24$. Thus in a network with more than 24 terminals, it pays to voluntarily limit the transmission ranges so that on the average only N=4.99725 other terminals are within hearing range. Figure 2.7 demonstrates graphically the result we saw in the previous pages that increasing the capture-parameter improves system performance. Here we plot γ' as a function of p for a fixed N and various values of β. In Table 2.2 we have listed the maximum γ' over all possible N and p values for a fixed β value. We note again that the γ' is increasing in β. Observe that the spread of the optimal values of N and p over all β values is not wide. Since these values do not change substantially with β, the capture-parameter of packet radios in the network does not need to be known to a high degree of accuracy to determine the network's N and p values that achieve optimal performance. In Table 2.2 we have also listed the probability of a successful transmission as well as the expected forward progress for the same N and p values. By comparing Tables 2.1 and 2.2, we observe that values of N and p which maximize γ' do not also maximize $P[E_s]$. Although it might seem intuitive that maximizing the number of successes in the network by picking an optimal transmission range R and hence by picking $N = \lambda\pi R^2$ would increase the throughput of the system, a little thought shows that this is not necessarily true. We can see this from Table 2.1 where the N values that maximize $P[E_s]$ are seen to be small, approximately 2.6 for $\beta = .7$. The network is in this case divided into many receiver-transmitter pairs in an attempt to take full advantage of the spatial reuse of the channel and although this increases the probability of successful transmission, packets in such an environment must pass over many hops before reaching their final destinations. This tends to decrease the number of packets reaching their final destinations in any one slot, and thus reduces the throughput of the system. This tradeoff between the probability of success and the throughput of the system as governed by the number of hops between source and destination is a fundamental issue of multi-hop systems and occurs in several guises. For example, we have already shown that increasing β will increase γ'. This increase in γ' can result from an increased $P[E_s]$, an increased $\bar{r}$ (thus increasing the average

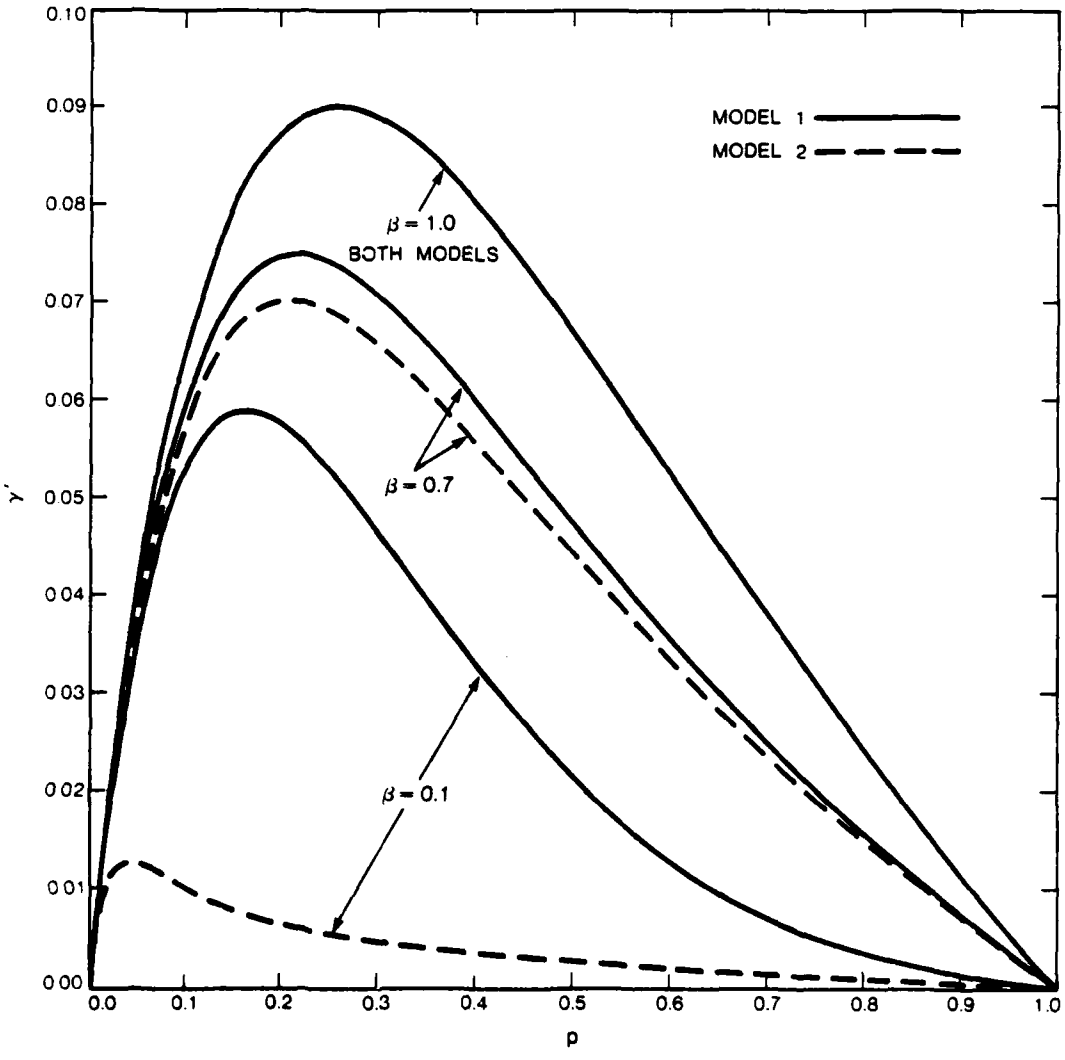


FIGURE 2.7
Normalized Throughput as a Function of p.
for $N = 5$ and $\beta = 0.1, 0.7$, and 1.0.

β	N	p	γ'	$P[E_s]$	$\bar{z}$
0.0	4.33261	0.18012	0.0584586	0.05991	0.42441
0.1	4.36181	0.18193	0.0591979	0.06285	0.40834
0.2	4.41670	0.18531	0.0605900	0.06577	0.39687
0.3	4.49194	0.18978	0.0624732	0.06879	0.38797
0.4	4.58656	0.19523	0.0648259	0.07190	0.38094
0.5	4.70160	0.20155	0.0676626	0.07526	0.37541
0.6	4.83818	0.20867	0.0710160	0.07875	0.37120
0.7	4.99725	0.21647	0.0749282	0.08242	0.36823
0.8	5.17892	0.22474	0.0794432	0.08624	0.36649
0.9	5.38063	0.23324	0.0845999	0.09022	0.36601
1.0	5.59807	0.24164	0.0904239	0.09433	0.36682

TABLE 2.2

number of hops a packet takes from source to destination), or a combination of both. We see in Table 2.2 that as β increases from 0 to .9, $P[E_s]$ increases and $\bar{z}$ decreases. Thus for optimal throughput packets must travel over more hops but they "hop" more frequently, once again showing the tradeoff between $P[E_s]$ and $\bar{z}$.

In Figure 2.8 we show the relationship of γ' as a function of p for fixed N and β. We notice that for any N, optimal performance is degraded for small changes of p from its optimal value, p^*, but that as N increases, the curves around this p^* become narrow. This variation of γ' for large N results from the fact that the transmission of any packet radio interferes with a larger number of other terminals. This increase in the number of collisions increases the sensitivity of the throughput for perturbations of the transmission probability from its optimal value. We also observe, for distinct values N_1 and N_2, such that $N_1 > N_2$, there exists a transmission probability, p^* such that for all $p_1 < p^* < p_2$ then $\gamma'(N_2,p_1,\beta) > \gamma'(N_1,p_1,\beta)$ and $\gamma'(N_2,p_2,\beta) < \gamma'(N_1,p_2,\beta)$. This can best be seen by looking at Figure 2.8 where we see that for the two model 1 curves for N=2 and N=5, there exist $p^* \approx .45$ such that, for all $p > p^*$ the N=2 curve dominates (has a greater throughput) the N=5 curve, and the opposite holds for $p > p^*$. Intuitively, we expect that for large p, systems with small N would have better performance characteristics since being less densely populated they are inherently more immune to interference than systems with higher N values. On the other hand for small p, since the time between transmissions grows larger in the small N case, one would expect that systems with higher N would have better performance.

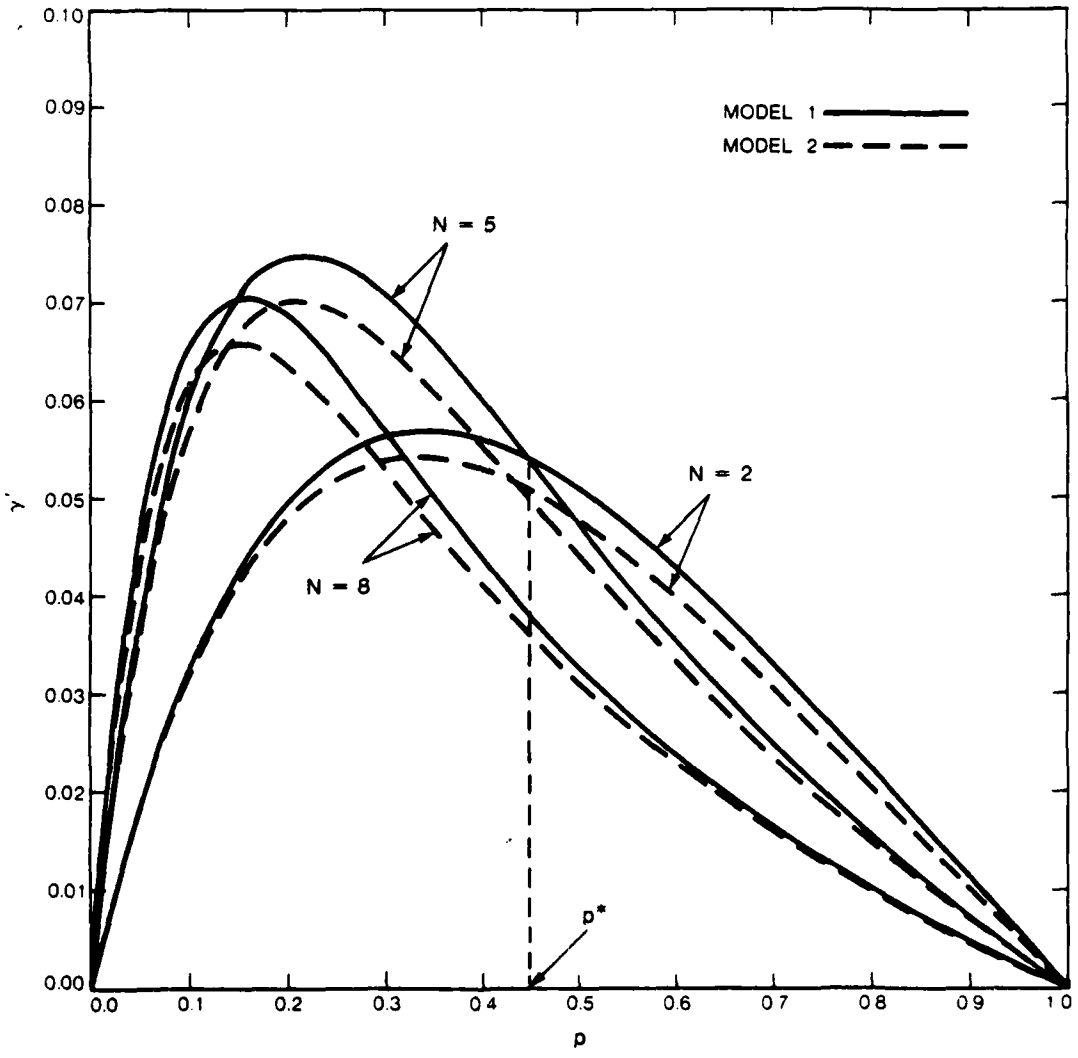


FIGURE 2.8
Normalized Throughput as a Function of p.
for $\beta = 0.7$ and $N = 2, 5$, and 8.

We can unify our discussion of Figure 2.8 by defining the offered load per unit area to be $G = Np$. From previous results [Abra73a] for finite population slotted-ALOHA networks, we know that $G = 1$ optimizes network throughput. In the multi-hop environment however, connectivity of the network must be preserved. This consideration, as noted before, manifests itself in the $c = (1-e^{-N/2})$ term appearing in the equation for γ. If p is large, $G = 1$ implies that $N = 1/p \approx 1$ which tends to disconnect the network since $c \approx .39$. Obviously optimal throughput for this case would have $N > 1$ and thus $G > 1$. We would thus expect $G = 1$ only in cases where p is small enough to make N sufficiently large to assure connectivity. We must however take account of capture in discussing the offered load. In the non-capture environment we would expect the offered load that maximizes throughput to be less than that for the capture environment because the probability a transmission suffers a collision is greater for $\beta = 0$ than for $\beta = 1$. Thus increasing $G = Np$ has a greater effect in increasing the number of expected collisions in environments with non-capture than for those with perfect capture. To check this intuition, we numerically calculated the N value that maximized γ' for fixed β and p and plotted $G = Np$ against p in Figure 2.9. We see that curves for high β values dominate those for lesser values, justifying our belief that the offered load that maximizes throughput can be greater for larger β, and that as p increases so does G, illustrating the relationship of the connectivity factor c has in sparse environments. We can lend some mathematical insight into these graphs by defining the *effective number of neighbors*, N', to satisfy $N = N'/(1 - e^{-N'/2})$. One then interprets N as the the average number of terminals per unit area that results when we randomly distribute terminals on the plane with an average density of N' terminals per unit area, and condition upon having a connected network. The $(1 - e^{-N'/2})$ term is thus seen to be the conditional probability of hop connectivity. Using N' instead of N in calculating the offered load, $G' = p\,N' \approx 1$ implies that $N' \approx 1/p$ and using this in the above equation that defined N', we would have $N \approx \hat{N} = 1/p(1 - e^{-1/2p})$. To check this intuition, we used the values obtained in generating the offered load curves of Figure 2.9 for $\beta = 1$ where we have seen that for low p values, $G \approx 1$. In Table 2.3 we produce the N and p values that optimize γ', as well as their product, the offered load G in the left part of the Table. In the right we tabulate the hypothesized values using the effective number of neighbors N'. We see that N and $\hat{N}$ are approximately equal and that the effective number of neighbors, N' is strictly less that N. Comparing G and G' we see that G' is much closer to 1 throughout the range of p lending support to our previous intuitive arguments.

In our last plot for this section, Figure 2.10, we graph γ as a function of N for $\beta = .7$ and various values of p. Observe that for the near optimal N for $p = .2$, (namely $N = 5$) the curve is very flat. This implies it is not necessary to determine N to a high degree of accuracy to achieve near optimal performance.

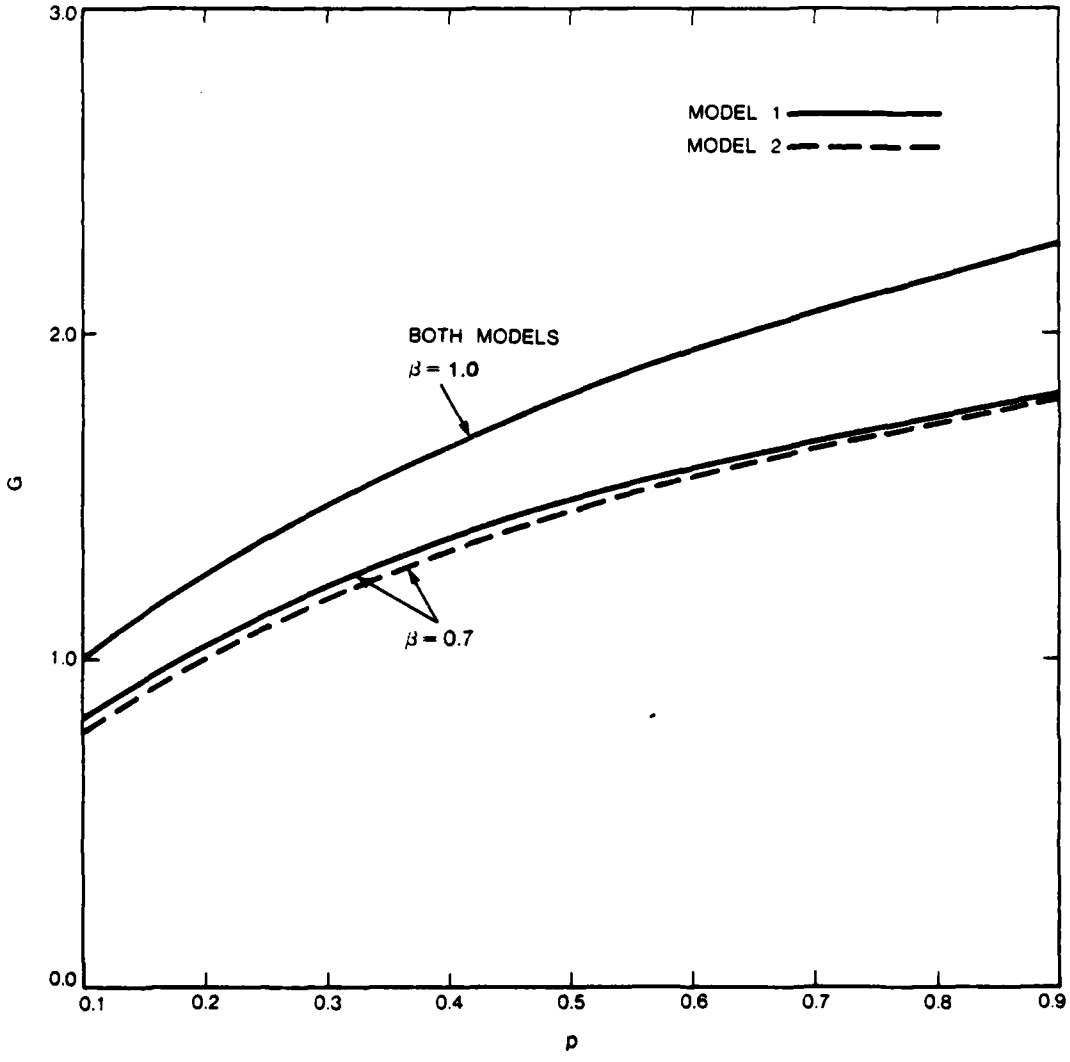


FIGURE 2.9
$G = Np$ as a Function of p
for optimal N and p, and $\beta = 0.7$ and 1.0.

N	p	$G = Np$	$\hat{N} \approx N$	N'	$G' \approx N'p$
10.07969	0.1	1.0079	10.0678	10.0121	1.0012
6.28435	0.2	1.2568	5.4471	5.9661	1.1932
4.91683	0.3	1.4750	4.1095	4.3614	1.3084
4.14327	0.4	1.6573	3.5038	3.3779	1.3511
3.62592	0.5	1.8129	3.1639	2.6733	1.3366
3.24784	0.6	1.9487	2.9477	2.1259	1.2755
2.95583	0.7	2.0690	2.7986	1.6793	1.1755
2.72158	0.8	2.1772	2.6896	1.3027	1.0421
2.52836	0.9	2.2755	2.6067	0.9774	0.8796

TABLE 2.3

2.4 Analysis of Model 2

Recall in Model 2 we assume that any other transmitter within $r\beta^{-1/2}$ of a packet radio receiving a packet from another transmitter a distance r away, will cause a collision on the channel. We thus do not need to divide the clean area into two regions. We will be brief in describing the results of this section, since most derivations follow lines similar to those in model 1.

2.4.1 Expected Number of Successful Receptions

The clean area for a transmitter at a distance of r from the receiver is now $r\beta^{-1/2}$, and thus the probability of this area has no other transmitters is $e^{-\lambda p \pi r^2/\beta}$. we thus have:

$$p[E_s] = p\ (1-p)(1-e^{-N/2}) \int_0^R \frac{2\ r}{R^2}\ e^{-\lambda p \pi r^2/\beta}\ dr$$

Integrating this over $[0,R]$ yields:

$$P[E_s] = \frac{\beta\ (1-p)(1-e^{-N/2})}{N}\ (1-e^{-Np/\beta})$$

and hence the number of successful receptions is

$$E = \frac{n}{N}\ \beta\ (1-p)(1-e^{-N/2})\ (1-e^{-Np/\beta})$$

Once again, the expected number of successes in the network is an increasing function in β.

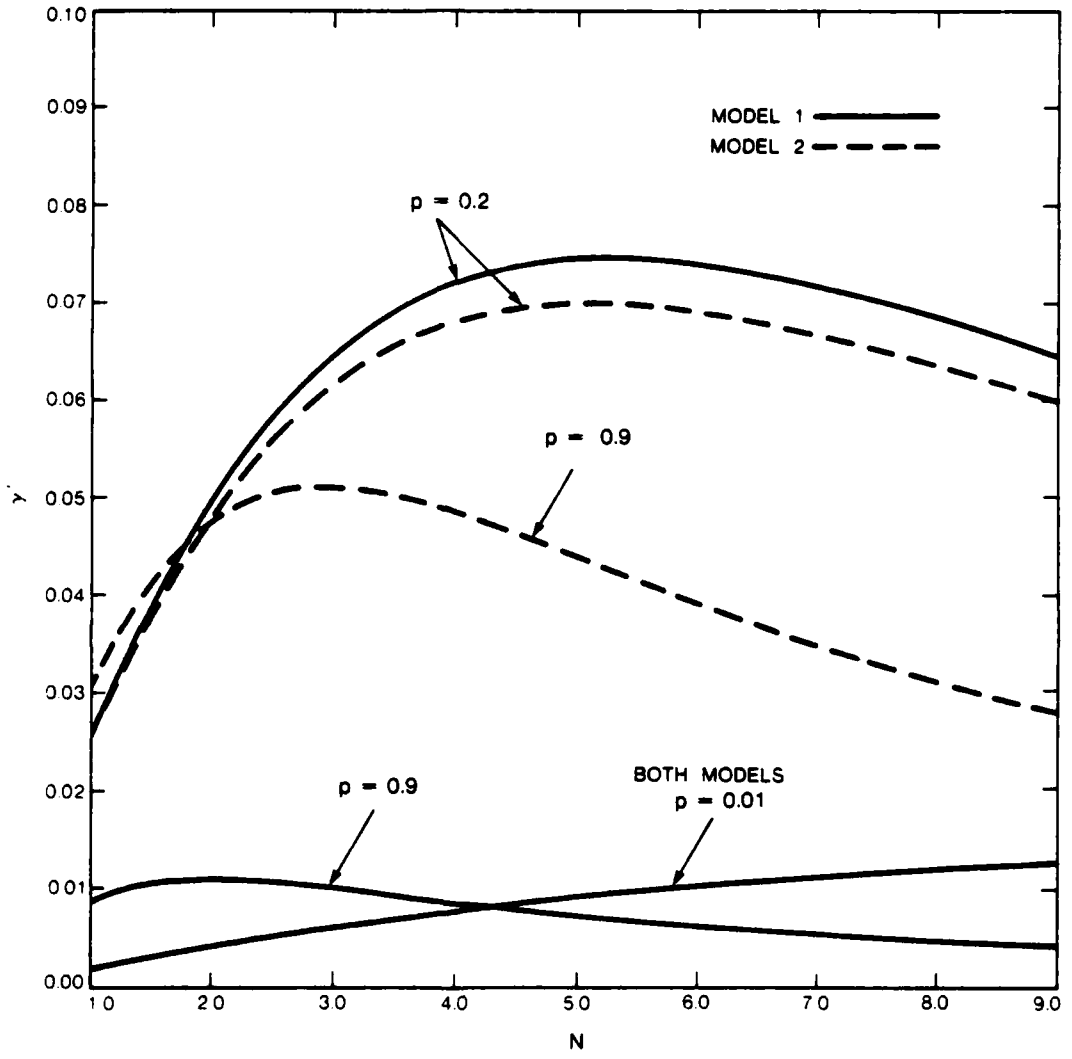


FIGURE 2.10
Normalized Throughput as a Function of N
for $\beta = 0.7$ and $p = 0.01, 0.2$, and 0.9.

This can best be seen by writing E as a function of β and showing that the slope is positive. Let $H(\beta) = K\,\beta(1-e^{-Np/\beta})$ where $K = n(1-p)(1-e^{-N/2})/N$. Differentiating $H(\beta)$ with respect to β yields $H'(\beta) = 1 - e^{-Np/\beta}(Np/\beta + 1)$ which must be greater than zero since $e^{Np/\beta} > 1 + Np/\beta$. It makes no sense to try to obtain the limit of the above expression for the infinite population, Poisson traffic slotted-ALOHA model as we did in model 1. The reason for this concerns our definition of capture for model 2. If we let β go to zero, the capture radius $r\,\beta^{-1/2}$ goes to infinity. Since all our derivations assume terminals to be Poissonly distributed over the plane with parameter λ, this infinite capture radius will contain an infinity of terminals which for any $p > 0$ will have an infinite number of transmitters with probability 1, thus guaranteeing a certain collision. The equation above rightly indicates that for $\beta = 0$ the expected number of successes is zero. In model 1, since the capture radius was limited to a maximum of R, it did make sense to let $\beta \to 0$ since the capture radius was bounded. Once again we can determine analytically that $E < n/2$ and produce Table 2.4 which contains the probability of success for various β values. Observe that for $\beta = 1$ these results agree with those of model 1.

β	N	p	$P[E_s]$
0.1	1.1295	0.20379	0.02737
0.2	1.5243	0.24990	0.04467
0.3	1.8104	0.27811	0.05794
0.4	2.0412	0.29824	0.06775
0.5	2.2373	0.31373	0.07789
0.6	2.4089	0.32622	0.08578
0.7	2.5621	0.33660	0.09272
0.8	2.7011	0.34543	0.09889
0.9	2.8284	0.35307	0.10443
1.0	2.9462	0.35977	0.10946

TABLE 2.4

2.4.2 Expected Forward Progress

We first calculate the density for the distance between a successful transmitter and its receiver. Using the same definitions as in the previous section, we have

$$g(r) = \frac{p\,(1-p)(1-e^{-N/2})}{P[E_s]}\,\frac{2\,r}{R^2}\,e^{-\lambda p\pi r^2/\beta}$$

defining $\frac{1}{s} = \frac{p\,(1-p)(1-e^{-N/2})}{P[E_s]}$ we can write $g(r)$ as

$$g(r) = \frac{1}{s}\,\frac{2\,r}{R^2}\,e^{-\lambda p \pi r^2/\beta}$$

We thus calculate the expected forward progress for each transmission to be

$$E[Z] = \frac{4}{s\,\pi\,R^2}\int_0^R r^2\,e^{-kr^2}\,dr$$

where $k = \lambda\pi p/\beta$. Using the result from Appendix A to expand the integral and then simplifying, we finally obtain

$$\bar{z} = \frac{2\,R}{\pi(1-e^{-Np/\beta})}\,e^{-Np/\beta}\sum_{j=1}^{\infty}(4Np/\beta)^j\frac{j!}{(2j+1)!} \tag{2.9}$$

2.4.3 Expected Throughput

Using the previous results, we can now write

$$\gamma = \frac{45}{64}\sqrt{n/N}\;\beta\,(1-p)(1-e^{-N/2})\,e^{-Np/\beta}\sum_{j=1}^{\infty}(4Np/\beta)^j\frac{j!}{(2j+1)!} \tag{2.10}$$

Again this is an increasing function of β, has the square root dependency on the number of nodes in the network, and explicitly accounts for $p = 0$, $p = 1$ cases and network connectivity.

2.4.4 Discussion of Results

To compare the performance of the two models we produced the same Tables and graphs for Model 2 as we did for Model 1. Earlier we observed that the two models should be identical for $\beta = 1$ since the clean area in both models for this case are identical. Algebraic comparison between similar formulas from both models shows that they are equal for $\beta = 1$ and the results in this section show that as β grows larger, the results of the two models are approximately equal. We therefore restrict our discussion of the curves and Tables in this section to relevant differences between the two models.

β	N	p	γ'	$P[E_s]$	$\bar{\tau}$
0.1	3.02345	0.06747	0.0136244	0.02092	0.33920
0.2	3.42712	0.10981	0.0254929	0.03610	0.34538
0.3	3.77498	0.14025	0.0360910	0.04805	0.35002
0.4	4.08648	0.16375	0.0457039	0.05787	0.35371
0.5	4.37335	0.18264	0.0545213	0.06616	0.35676
0.6	4.64158	0.19828	0.0626785	0.07330	0.35936
0.7	4.89561	0.21153	0.0702766	0.07953	0.36159
0.8	5.13830	0.22293	0.0773940	0.08503	0.36355
0.9	5.37173	0.23289	0.0840929	0.08993	0.36528
1.0	5.59807	0.24164	0.0904239	0.09433	0.36682

TABLE 2.5

We observe in Table 2.5 that for model 2 there is a wider spread for optimum N and p values for the range of β values. This is due to the larger capture radius for small β. One curious difference between this and Table 2.2 is that here as β increases so does $\bar{\tau}$, whereas in model 1 we observed a decrease in the $\bar{\tau}$ values. The increase for model 2 can be explained by the fact that for low values of β, small values of r, the distance between transmitter and receiver, have higher probability than greater r values because the capture radius for these values is large. As β increases, the capture radius for a fixed r, decreases in model 2 and thus larger r values are more heavily weighted and thus increase $\bar{\tau}$. This explains the increase in $\bar{\tau}$ for model 2. In model 1 however, the capture radius was bounded to be less than R. Thus the clean area is small for terminals that lie close to R and this increases the probability that the projected distance will be large. This is the reason why the throughput for model 1 is always larger than that for model 2.

In Figure 2.7 we plot γ as a function of p for $N = 5$ and various β values. Once again we see the dominance of the higher β curves. We can also observe that curves for model 1 dominate those for model 2. Comparison of the curves for the two models on this plot shows the dominance of throughput values of model 1 over those of model 2. This dominance is a result that the probability of collision is much larger in the second model for any value of N and p and hence decreases the expected throughput of the system. Figure 2.9 shows again the fluctuations of the offered load for increasing p values and we note that the curves for $\beta = 0.7$ is almost identical to the same curve for model 1. Figure 2.10 shows the flatness of throughput curves for fixed β and p, over values of N.

2.5 Conclusions

In this chapter we have analyzed two models of capture in a random planar network where slotted-ALOHA was used to broadcast packets on the channel. The results of the two models are similar for capture-ratios achievable on good FM receivers and thus either could be used to analyze networks of this kind. Since increasing the capture parameter increases the throughput of the network we can conclude that capture is a beneficial phenomenon and throughout the following chapters all of our network models will assume that radios can capture signals. The tradeoff between the probability of a successful transmission and the expected number of hops taken by a packet in the network has been delineated, and we have seen that even in ideal conditions with perfect capture and one-hop messages, no more than 21% of the nodes in the network on the average, are engaged in productive communications in each slot. The square root of n dependency on the throughput has been shown to substantially increase the throughput of the network over conventional one-hop ALOHA networks when the number of nodes in the network is sufficiently large. The critical parameter to network optimization has been shown to be the value of p, the probability of transmitting in any given slot and we have seen that an offered load of $G \approx 1$ optimizes network performance.

In the following chapter we eliminate the slotted-ALOHA assumption from this model and address a more basic question. In that chapter we seek to determine the throughput of the best protocol in the mobile multi-hop environment. We will derive a tight upper bound for the performance of an optimal protocol in this environment and then use that as a standard to determine how well slotted-ALOHA performs in comparison to this approximation. Some of the observations we have made in this chapter, we will see, generalize to this optimal protocol.

CHAPTER 3
AN UPPER BOUND FOR THROUGHPUT FOR ALL PROTOCOLS IN A MOBILE NETWORK

From Chapter 2 we know how well slotted-ALOHA performs in a multi-hop mobile packet radio network. In this chapter we remove the slotted-ALOHA assumption from the network model and determine an upper bound on the best performance for any protocol in a mobile multi-hop network.

3.1 Introduction

In the previous chapter we analyzed a packet radio network consisting of randomly distributed nodes on the plane and determined optimal range and transmission probabilities for the slotted-ALOHA protocol. One of the parameters of the system, N, the average number of neighboring nodes, contained the range factor (since $N=\lambda\pi R^2$, where λ was the density of nodes on the plane and R was the fixed range of transmission) and was varied to maximize the throughput of the network. In this chapter we assume we are given a connected, random, network with a fixed N (it is not a parameter that can vary) and seek to determine, under these assumptions, the maximum number of simultaneous transmissions in the network that can be successful. This number, divided by the number of nodes in the network, will give the maximum fraction of successful transmissions that are possible in networks of this type and will be an upper bound for any random access protocol. This can be used then as a standard to evaluate the performance of other protocols. In Chapter 4 we will refine the definition of an optimal protocol of this type and will derive its delay characteristics.

3.2 Model and Analysis

We will assume that nodes of a packet radio network are distributed on the plane according to a Poisson point process with a mean density of λ radio units per unit area. With no loss of generality we will assume that radios transmit with a range of one unit ($R=1$). The average number of neighbors they have, N, therefore, is given by $N=\lambda\pi$. From any given node, a, another node, b, is said to be i-hops away if there exists a path from a to b that contains $i-1$ other nodes and no other path exists between a and b that contains fewer nodes. In our derivation we will focus on a section of the network containing n nodes, and thus will be

concerned with a portion of the network of radius R_0, where $n = \lambda \pi R_0^2$. Our final results are independent of n. We will assume the network formed by these n nodes is connected. We should mention here that our network model, as in Chapter 2, is that of a mobile packet radio network (thus the random distribution of nodes on the plane). In such networks it is extremely difficult for nodes to ascertain the locations of other nodes in the network, and, in particular, nodes do not know the direction of the recipients of their transmitted packets. Our results here are applicable only to protocols that do not utilize information concerning the location or direction of their neighboring nodes.

To motivate how we propose to calculate our upper bound suppose a node, a, is transmitting a packet to one of its neighbors. Let S_i be the set of nodes that are i-hops away from a, and define a *k-order independent set* to be a set of nodes that are all mutually k or more hops away from each other. A *maximal k-order independent set* is a k-order independent set to which no other node of the network can be added. It is clear that if all nodes of a maximal 3-order independent set transmit, and these are the only transmitting nodes in the network, all of their transmissions will be *successfully received*, where by this we mean that nodes in the network that are possible recipients of any of these messages, hear exactly one transmitter.

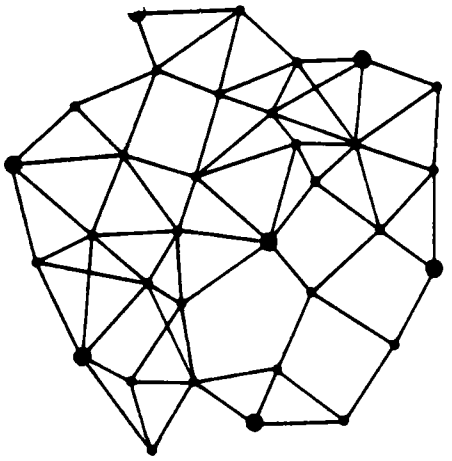


FIGURE 3.1
A maximal 3-order independent set.

An example of such a set is shown in figure 3.1. In this figure, nodes of the 3-order independent set are indicated by the larger circles. We can easily show the following property of maximal 3-order independent sets:

Lemma: For any maximal 3-order independent set M, no node not in the set can transmit without causing interference with a transmission of at least one node in M.

Proof: Suppose a was such a node. Using the previous definitions, it is clear that $S_1 \bigcap M = \varnothing$ (where $\varnothing$ is the null set) otherwise a would be interfering with its own reception from nodes in M. If $S_2 \bigcap M \neq \varnothing$ then this implies there is a node that simultaneously receives signals from a and also from another node in M. Thus $S_2 \bigcap M = \varnothing$, but this implies that M is not maximal since a is at least 3 or more hops away from every node in M.

■

We should observe that this lemma is also true for maximal k-order independent sets where $k = 4, 5$, and thus we cannot immediately conclude that a maximal 3-order independent set corresponds to the greatest number of transmissions in the network that are guaranteed to cause no collisions. Intuitively, however, to achieve maximal throughput we would want transmitters to be as close to each other as possible, without having collisions detract from the throughput of the channel. To make this more precise, let S_k be the set of all maximal k-order independent sets, and let L_k be the cardinality of the largest set in S_k (we are assuming here a finite but arbitrarily large graph). Then, since a $k+1$-order independent set is also a k-order independent set, we have that $L_k \geq L_{k+1}$. Thus, since collisions occur for $k = 1, 2$, we can conclude that the largest number of successful transmissions, without allowing collisions, is given by L_3.

Since nodes in this scenario that are allowed to transmit are mutually at least three hops away from each other and at best exactly three hops, in the ideal case we can imagine the plane being tessellated with equilateral triangles having sides equal to the average distance between nodes three hops away. This tessellation is motivated in figure 3.2 where we have connected the nodes of the maximal 3-order set of figure 3.1 (not all such configurations will result in a hexagonal shaped figure). For such a tessellation each vertex corresponds to a transmitting node of the network. The number of such triangles, for a given section of the network, will correspond to twice the number of vertices. This can be seen in figure 3.3 where we have mapped each vertex to the triangle lying directly above it (shown by the arrows in the figure). The shaded triangles have no corresponding vertices and can then be seen to be equal in number to those that are mapped to vertices. If we let $\bar{X}$ be the average distance between nodes three hops away, then the area of each triangle is given by $\sqrt{3}\,\bar{X}^2/4$. Letting T be the number of triangles in network under consideration whose total area equals πR_0^2, we can write:

$$T = \pi R_0^2 / (\sqrt{3}\bar{X}^2/4)$$

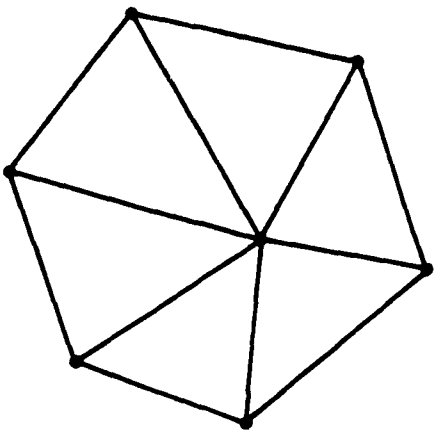


FIGURE 3.2
Tessellation of Figure 3.1.

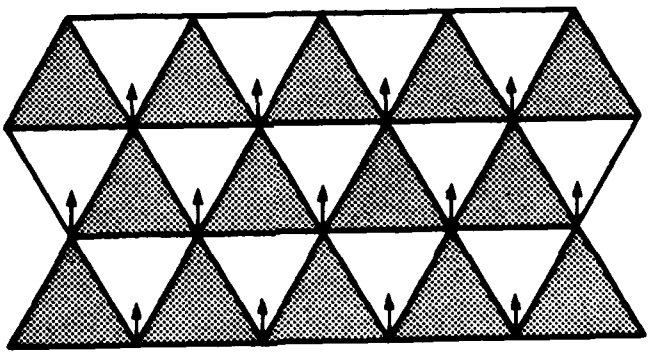


FIGURE 3.3
Mapping of vertices to triangles.

The fraction f of successful transmissions then can be written as:

$$f = T/2n = \frac{2}{\lambda\sqrt{3}\bar{X}^2}$$

In Appendix B it is shown that $\bar{X} = 2$ and thus we have (using the fact that $N = \lambda\pi$):

$$f(N) = .9068/N \tag{3.1}$$

This then represents the fraction of successful transmissions that occur in a network using a protocol that schedules transmissions in a manner that at all times a maximum number of nodes in the network transmit and there is no possibility for collisions. To check the intuition that lead to this equation we generated random planar connected graphs with different mean densities. These graphs had from 50 to 90 nodes. Using these graphs we found the size of the maximal 3-order independent set. The fraction of these nodes was then calculated. In Figure 3.4 we have plotted $f(N)$ as well as these generated values and we see a close match.

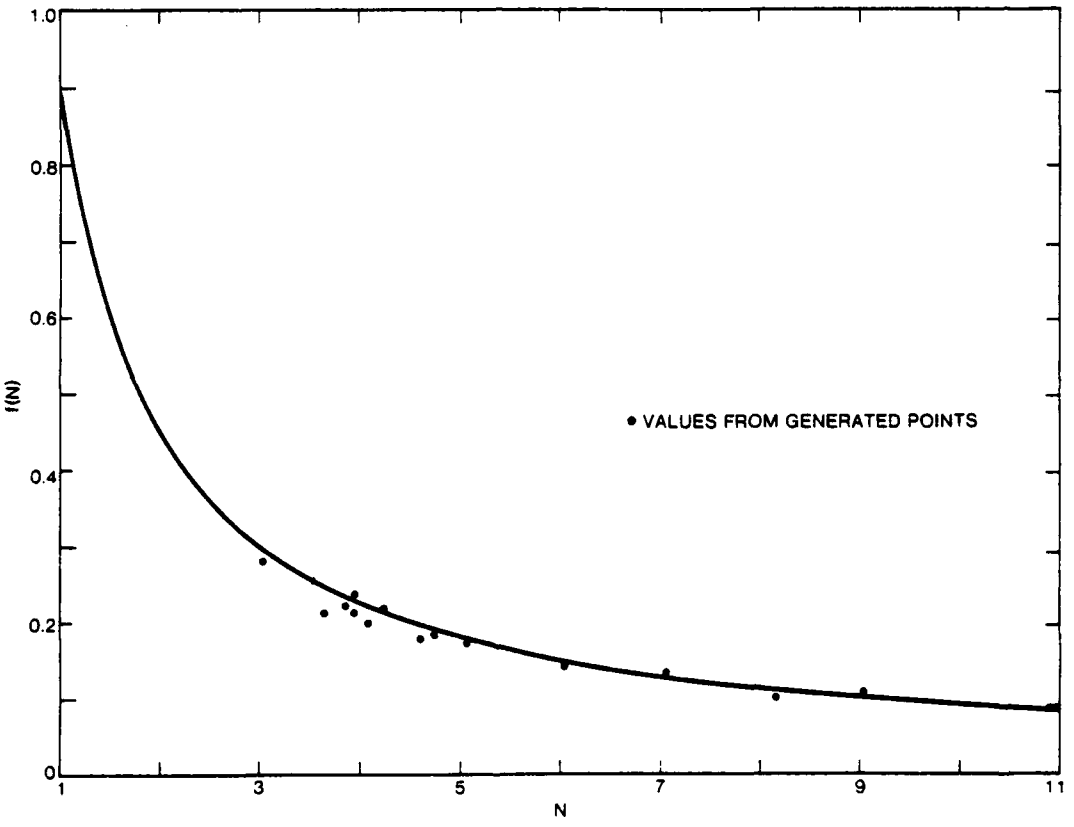


FIGURE 3.4
Comparing $f(N)$ with Generated Data.

The data that lies above the $f(N)$ curve is a result of the edge-effects obtained from generating finite graphs. In such graphs, nodes at the edge of the graph are more likely to belong to maximal independent sets since they have neighbors only on one side. This tends to increase the

size of the maximal 3-order independent sets.

It is interesting to graphically see what $\bar{X} = 2$ means in terms of transmission areas.

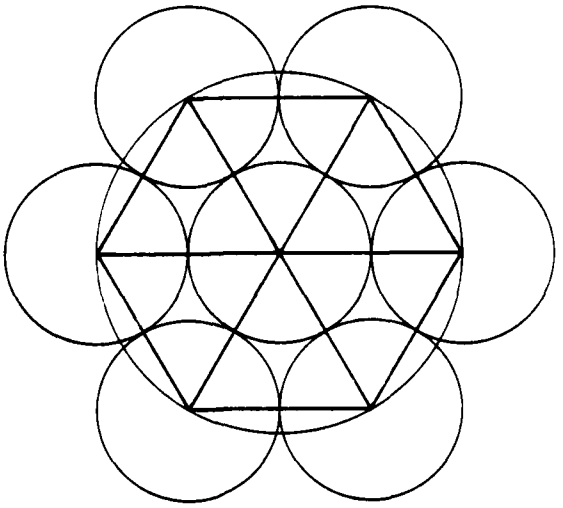


FIGURE 3.5
$\bar{X}$ in terms of transmission areas.

We see in figure 3.5 that this average third hop distance implies that the plane is covered with unit circles from transmitters located at the vertices of the equilateral triangles. Observe that there is no overlap of the circles and that very little of the plane is not covered. In fact it is well known that this arrangement of circles maximizes the density of the area of the plane covered by exactly one circle provided that no area is covered twice [Will79a]. Viewing the problem in this manner leads us to inquire about the arrangement, allowing possible overlap of circles, that yields the highest density of singly covered area. This density will yield an upper bound to the fraction of the expected number of successful transmissions for a mobile packet radio network since, in the absence of capture, the probability that a transmission is successfully received is equal to the fraction of area that is covered by only its transmission. We can use figure 3.6 to this end. This figure is a reproduction of one of the triangles of figure 3.5, where we have drawn the triangles with a length less than 2 units. In this figure let $I(x)$ be the area of overlap within the triangle, and let $E(x)$ be the area which is not covered in the circle tessellation. First let us derive equations for these two functions. We can use the equation derived in Appendix B for $A(\cdot)$ (equation (B.7)) to write:

$$I(x) = (\pi - A(x,1))/2$$

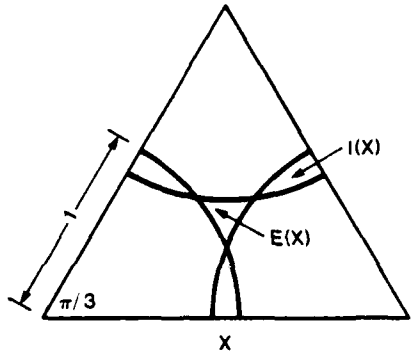


FIGURE 3.6
Section of Plane with Overlapped Tessellation.

which after some manipulation becomes:

$$I(x) = \cos^{-1}(x/2) - \frac{x}{2}\sqrt{1-\frac{x^2}{4}}$$

The amount of area within the triangle that is covered is $C(x) = \pi/2 - 3I(x)$ and subtracting this from the area of the triangle gives:

$$E(x) = \frac{\sqrt{3}}{4}x^2 - \frac{\pi}{2} + 3\cos^{-1}(x/2) - \frac{3x}{2}\sqrt{1-\frac{x^2}{4}}$$

Now suppose we wanted to "partially" tessellate the plane with unit circles in such a way that the sum of the overlapped and the uncovered areas was minimized (thus maximizing the amount of singly covered area). It is clear from figure 3.6, that $I(x)$ is increasing in x and $E(x)$ is decreasing, and thus seeking a minimum is a well formed problem. Letting the objective function be $F(x)$ we then have:

$$F(x) = \frac{\sqrt{3}}{4}x^2 - \frac{\pi}{2} + 6\cos^{-1}(x/2) - 3x\sqrt{1-\frac{x^2}{4}}$$

This function is minimized for $x^* = 1.9215$ which gives the tessellation shown in figure 3.7 (the points of intersection form a 12-gon). We can also derive the fraction of the plane that is covered by this type of tessellation by forming the ratio of the covered area to the area of the triangle. When this is done, the fraction is determined to be .9278, (as compared to .9068 of equation (3.1)) and thus about 92% of the plane is singly covered. We can thus write the

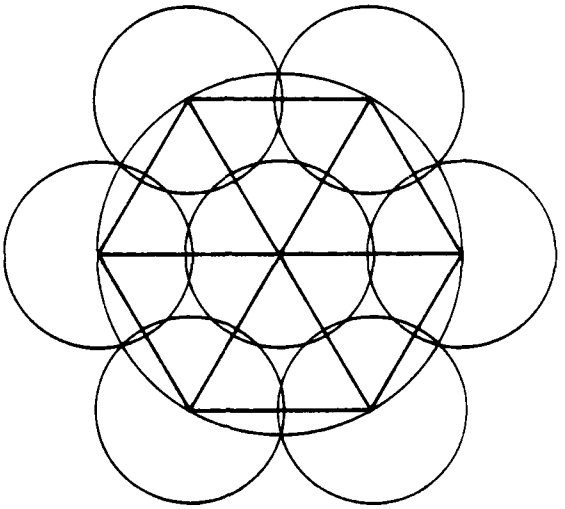


FIGURE 3.7
Allowing Collisions to Increase the Bound.

upper bound as:

$$f^*(N) = .9278/N \tag{3.2}$$

This arrangement of circles on the plane was conjectured to have the highest density of singly covered area in 1964 [Toth64a] and has recently been proved by Ernst Strass at UCLA (unpublished manuscript). We should observe that for reasonably large N, equations (3.1) and (3.2) agree for practical purposes and thus all the approximations and conclusions made using equation (3.1) are also applicable to our new bound.

If we let G be the offered load (the average number of packets presented to the channel per unit time) of traffic coming from the area of one circle, we clearly see that $G \cong 1$. This corroborates our results of the previous chapter where we found that $G \cong 1$ for the slotted-ALOHA network analyzed there. As mentioned in Chapter 2, this result is a generalization of that for a single-hop network where it can be shown that an offered load of one packet per unit time optimizes performance [Abra70a]. We can derive a rule of thumb for multi-hop packet radio networks by approximating equation (3.2) as:

$$f(N) \cong 1/N$$

It can best be interpreted by looking at figure 3.7. Here we see that the plane can also be viewed as being almost tessellated by circles of unit radius, the center of which contains a

transmitter. Since each transmitter sends to an average of N other nodes, it is clear that only one of N nodes will be successful, thus giving the approximate $1/N$ ratio.

3.3 Comparison to slotted-ALOHA

We can compare this optimal fraction of transmissions to that of the slotted-ALOHA radio network studied in the previous chapter. In that chapter, the equation for the probability of successful transmission in the non-capture case, (which corresponds to our $f^*(N)$) was given by:

$$P(N,p) = (1-p)p(1-e^{-N/2})e^{-Np} \tag{3.3}$$

where p is the probability of transmitting in any randomly selected slot. Recall the $(1-e^{-N/2})$ term in this equation is a factor that represents the probability that the network is connected. Since our derivation of $f^*(N)$ assumes a connected graph, we should eliminate the connectivity term from equation (3.3) before comparing the performance of this system to that of the optimal protocol.

For a given N, $P(N,p)$ achieves a maximal value for:

$$p^* \triangleq q(N) = \frac{N+2-\sqrt{N^2+4}}{2N}$$

We can thus write the maximum fraction of successful transmissions for the slotted-ALOHA network as:

$$P^*(N) = q(N)(1-q(N))e^{-Nq(N)} \tag{3.4}$$

The *efficiency* of a protocol is defined to be the ratio of its performance to that of the optimal protocol and thus we define:

$$e(N) \triangleq P^*(N)/f^*(N)$$

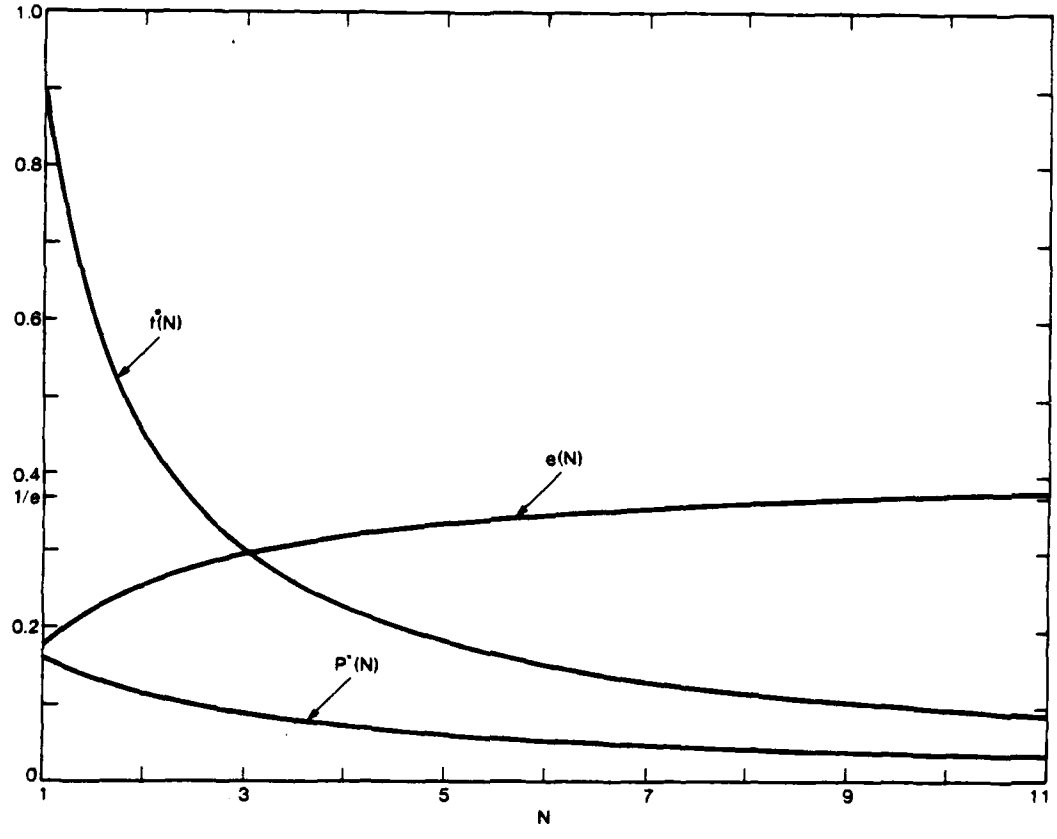


FIGURE 3.8
The Efficiency of slotted-ALOHA.

In figure 3.8 we have plotted $f^*(N)$, $P^*(N)$ and $e(N)$ for $1 \leqslant N \leqslant 11$. In this figure we observe that $e(N)$ increases in N. To find its limit, we observe:

$$\lim_{N \to \infty} q(N) = 1/N$$

hence we have:

$$\lim_{N \to \infty} e(N) = 1/(.9278\, e) = .396$$

Thus the capacity of slotted-ALOHA at best is about 40% of the channel capacity as given by our approximation. Observe that the optimal capacity of slotted-ALOHA occurs when the density of terminals approaches infinity. This of course corresponds to an unrealistic network configuration. The increase in efficiency as N grows larger is, however, very slow. For realistic size networks, looking at figure 3.8, we see that we can approximate the efficiency to be about

$1/e$. In a single-hop environment slotted-ALOHA achieves a maximum capacity of $1/e$ and thus we have the intuitively pleasing result that the maximum capacity of the slotted-ALOHA protocol is about $1/e$ in both single and multi-hop environments.

3.4 Comparison with CSMA

In this section we will use the method defined in the previous sections to derive an equation for the optimal performance for CSMA in a multi-hop environment and then will calculate the efficiency of this protocol. In a CSMA environment, nodes that hear an idle channel can transmit packets on the channel. In a random connected network the maximal number of nodes that could transmit using the CSMA protocol is equal to the size of the maximal 2-order independent set. We can calculate the fraction of nodes in such a network using methods developed in previous sections. If we imagine, in figure 3.3, the length of the equilateral triangle to be the average euclidean distance for nodes separated by 2 hops, defined to be $\overline{Y}$, then $g(N)$, the fraction of nodes in a maximal 2-order independent set, can be calculated as:

$$g(N) = 2.214/N$$

where we have used the result derived in Appendix B that $\overline{Y} = 1.2881$. If we assume the fraction of the nodes in a maximal 2-order independent set that transmit is p, then the fraction of transmissions in the network is given by $g(N)p$. Unlike the derivation of the optimal protocol however, in this case not all transmissions are successful. We can calculate the probability that a randomly selected transmitter is successful by using figure 3.9. In this figure we have shown the triangle tessellation corresponding to a maximal 2-order independent set, and the circular tessellation that results when nodes on the vertices of the triangles transmit. If we suppose in this figure that P is a transmitter, then we can calculate the probability that it is successful by determining the probability that P is sending its message to a neighboring node located in one of the areas labeled A, B, or C of the figure. We will assume a uniform probability in making this calculation and also assume that each transmission corresponds to a series of messages having infinitesimal length. This assumption allows us to calculate the expected fraction of messages that are successfully received before being collided by an interfering transmission. An alternative way to model this is to assume that each receiver captures the first signal that is sent to it. Although such *perfect time capture* [Davi80a] is not attainable in real networks, this assumption will allow us to calculate an upper bound for network performance. Area $C + B$ forms half of the intersection of two unit circles separated by a distance of $\overline{Y}$, and can be shown to be equal to $\pi/2 - A(\overline{Y}, 1)/2 = .3845$ where $A(\cdot)$ is derived in Appendix B (equation (B.7)). The area of the triangle is easily seen to be .7094 and thus we can write the following equations:

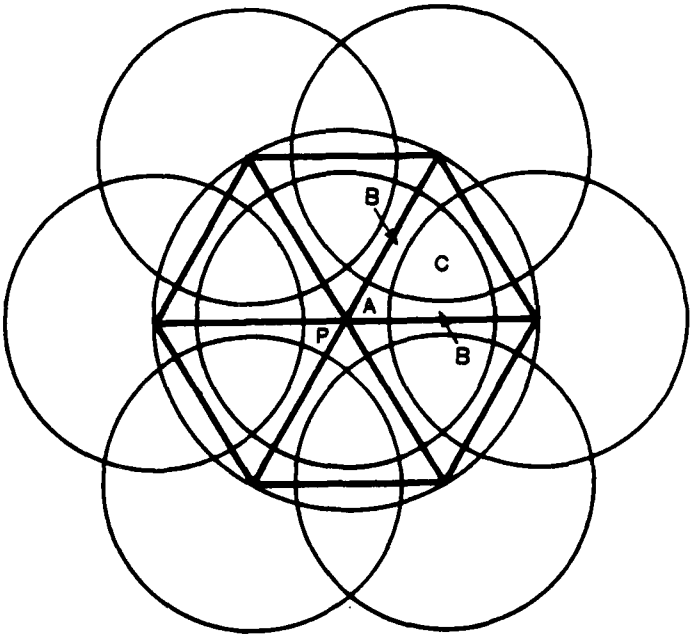


FIGURE 3.9
Tessellation for CSMA Calculation.

$$\begin{aligned} \pi/6 - .3845 &= A + B \\ .7094 &= 3(A+B) + C \end{aligned}$$

which can be solved to yield $A = .0467$, $B = .0923$, and $C = .2921$. We can convert these areas into probabilities by dividing by $\pi/6$ to obtain $P_A = .0891$, $P_B = .1762$, and $P_C = .5579$. Each area is influenced by only a subset of the possible transmitters surrounding it, and we can write the probability that P's transmission is successfully received as:

$$Q = P_A + 2P_B(1-p) + P_C(1-p)^2$$

The fraction of successful traffic then is given by $C(N,p) = g(N)pQ$ which is maximized for $p^* = .4599$ at which point $Q = .4421$. Using this in the expression for $C(N,p)$ allows us to calculate the maximal fraction of successful transmissions in the CSMA network which is given by:

$$C^*(N) = .4504/N$$

Thus we have the efficiency of the CSMA protocol in the multi-hop environment under ideal assumptions is about 48.5%. This performance is in striking contrast to the efficiency of CSMA in the single-hop environment [Klei75a] where under commonly held assumptions it has a

throughput of about 87%. We may also note that the performance of CSMA in the multi-hop environment under ideal assumptions is not much greater than that of slotted-ALOHA.

3.5 Densities of Θ, R_1, R_2 and R_3

As a by-product of this research, we derive in Appendix B the probability density functions for R_1, R_2, R_3, and Θ which are plotted in figures 3.10, 3.11, and 3.12. In table 1 we give the means and variances for these random variables.

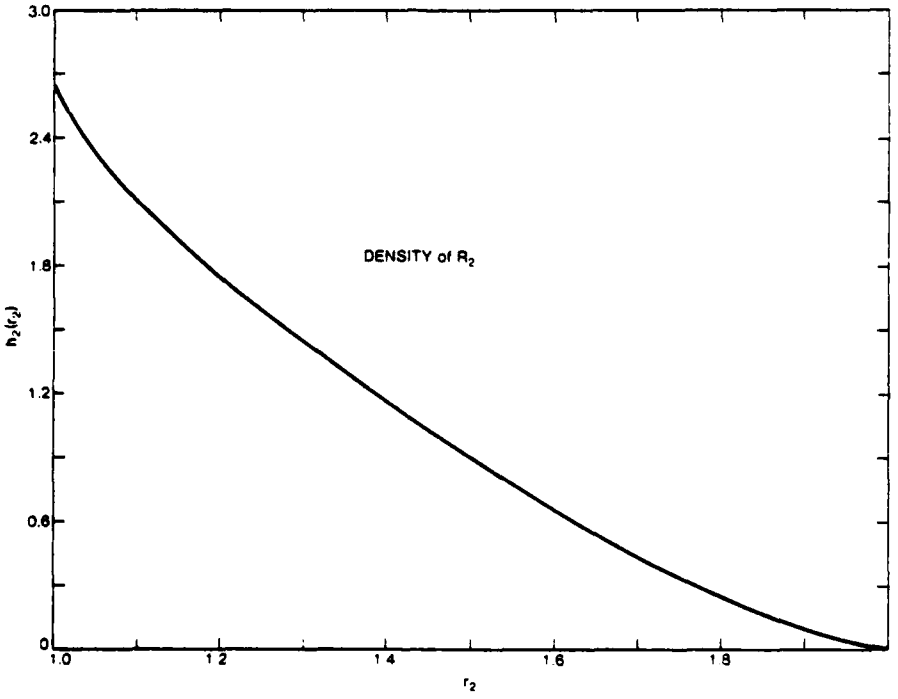


FIGURE 3.10
The Density of R_2.

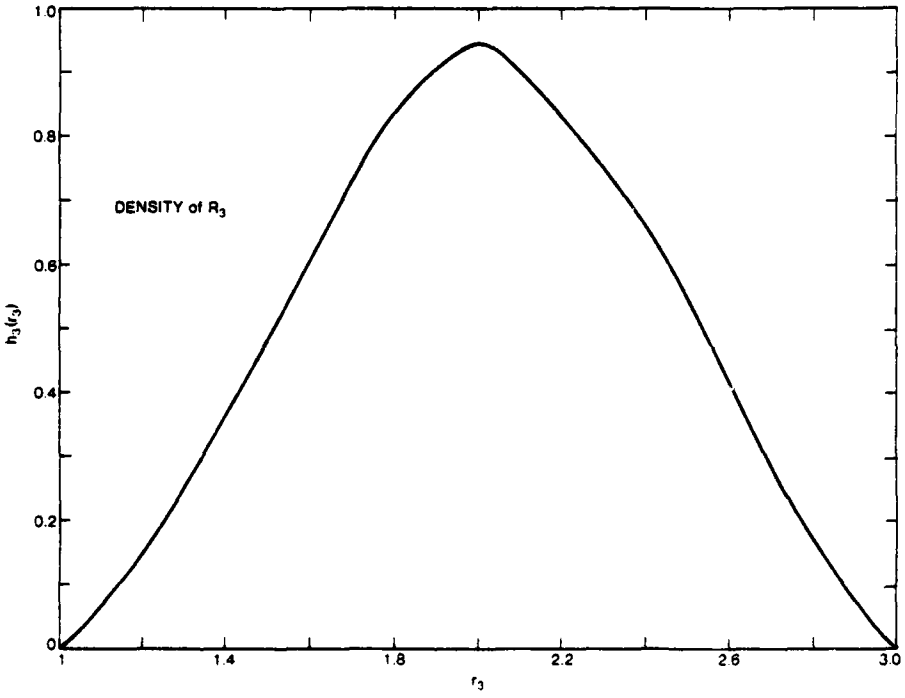


FIGURE 3.11
The Density of R_3.

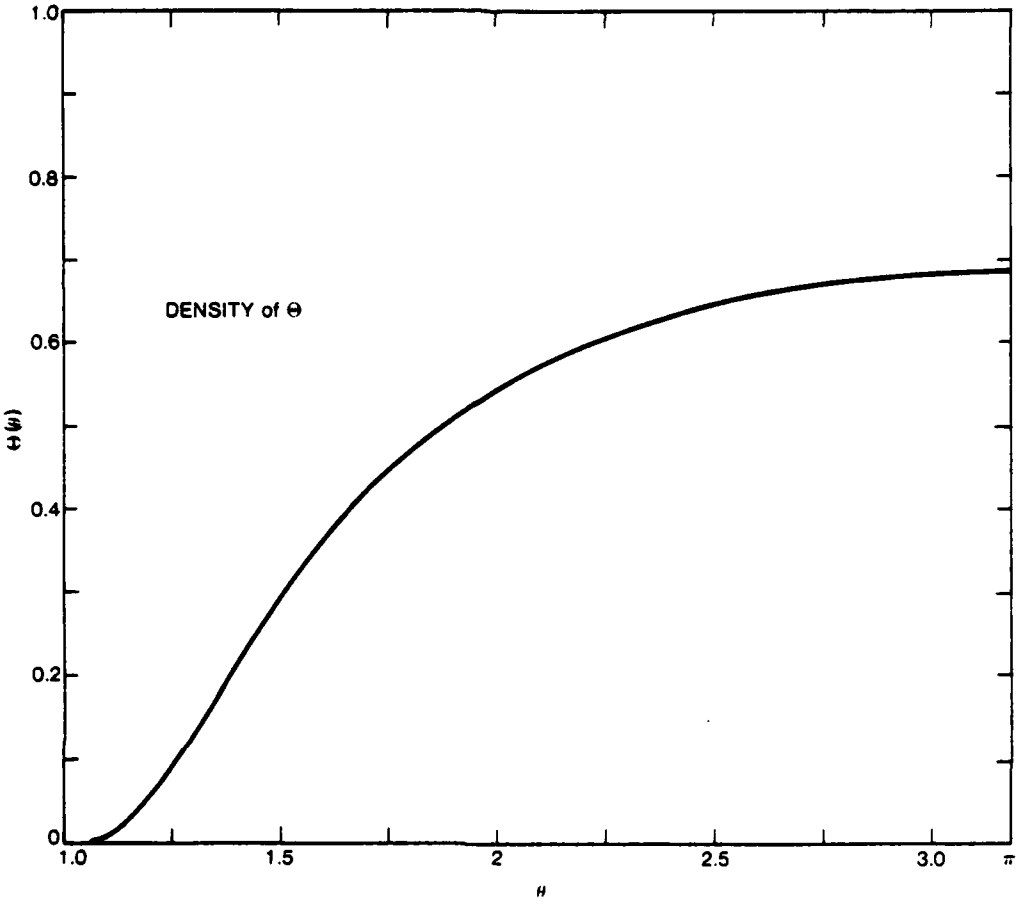


FIGURE 3.12
The Density of Θ.

Density	Mean	Variance
R_1	0.6667	0.0555
R_2	1.2881	0.0639
R_3	2.0000	0.1870
Θ	2.3254	0.2807

Table 1
Means and Variances for R_1, R_2, R_2, and Θ

3.6 Conclusions

In this chapter we have calculated an upper bound for the maximal expected fraction of successful transmissions obtained in a random connected planar packet radio network. The method utilized is notable in that it produced an intuitively pleasing result that $f(N) \cong 1/N$ and was simple, although tedious, to calculate. Using this optimal performance as a standard, we compared the slotted-ALOHA and CSMA protocols to it, and have shown that these protocols, for realistic networks, have an efficiency of about 36%, and 48.5% respectively. The probability density functions for the euclidean distance for nodes separated by one, two, and three hops was derived as a by-product of this research. In the next chapter we will investigate the delay characteristics of the optimal protocol which was discussed in this chapter. This will lead us to formulate a capacity assignment problem for networks using this protocol.

CHAPTER 4
SPATIAL—TDMA

In the previous chapter we determined the maximal throughput for a network which used an protocol that gave a maximum number of nodes in the network permission to transmit over all periods of time in a manner that did not allow collisions. In this chapter we formally define this protocol, called *spatial-TDMA*, and determine its delay characteristics. This leads us to formulate an optimal capacity assignment problem for networks of this type which we address in a later section. First we will review the motivation for defining such a protocol and the definition of the capacity assignment problem.

4.1 Introduction

Networks have traditionally consisted of a set of switches connected together by some form of cable. Because cable must be strung point to point, and often require land acquisition and construction of supporting structures, the design of such networks requires careful, long-range planning and it is not surprising that considerable research has been devoted to the problems of this initial design. One problem that has been extensively studied, for example arises when the locations of the nodes of the network and their traffic characteristics are assumed to be known. In this case, one must determine a procedure for choosing the capacities of the communication lines. Since the cost of the network is an increasing function of these capacities, it is important that the designer of such a network select the assignment of the capacity of the links of the network to minimize network expense (under the constraints of preserving a tolerable delay for messages in the system). Various solutions [Klei64a, Cant74a, Fran71a, Meis71a] have been proposed for this, and other problems of a similar genre, which are commonly classified as *Capacity Assignment Problems*. As one can expect however, network specifications often change, and an optimal design for an initial network may be far from optimal after changes are made either to the traffic characteristics of the nodes or to the network's topological structure. Indeed, even the minor change of optimally adding one new node to a wire network can be a formidable problem. Besides the difficulty of connecting cable from the new node to its adjacent neighbors, one must also solve the capacity assignment problem again for the changed network. If the traffic offered by the new node changes the loads on the links of the already existing network, then some of them will need to

be upgraded to higher capacities. Likewise, other lines might be able to have their capacities reduced if some of the flow through them can be routed over the new lines. Upgrading existing lines however, is expensive, and the cost of preserving this optimality in the network after the addition of a new node could be so formidable that the designers would have to settle for a less than optimal solution. As the process of adding new nodes to the net continued, it would not be surprising to see performance seriously deteriorate.

To avoid these problems, one needs to use a more flexible medium for inter-connecting the switches of the network. As we saw in Chapter 1, a broadcast medium such as radio, offers flexibility to topological changes, and has the property that capacities can be changed to reflect alterations in the specifications of the network. Connectivity in such networks is determined by the power of the transmitters of the network and thus is easily controlled, and the addition of new nodes is free of the cost of stringing cable. There still remains of course, the problem of re-adjusting the capacities of the lines of the system in an optimal manner. Indeed, determining the capacity of a specific link in a radio net, and adjusting the capacities of all the nodes in the network in some optimal manner, are not straightforward problems and depend upon the protocol that nodes use to determine when to transmit messages.

We address the above issues in this chapter and formulate a collision-free channel access protocol that nodes use to determine when they have permission to transmit. With this collision-free property, a multi-hop packet radio network can be made to *simulate* a wire network, and the determination of the capacity of an arc is straightforward. The basic idea of this protocol is to allocate specific periods, from a specified cycle time, during which nodes can transmit. This allocation has the property that during any period, no link is denied transmission rights if it would not potentially cause a collision (thus this is similar to the protocol that was discussed in Chapter 3). The capacity of an arc is then proportional to the amount of time from the cycle during which it has permission to transmit. We first will discuss the delay characteristics of such a protocol and then propose a method for assigning periods and adjusting their durations.

4.2 Description of the Protocol

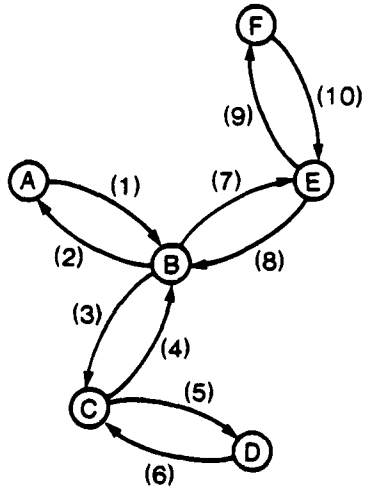
Throughout our discussion we will assume, as in wire networks, that the locations and connectivity of the packet radio transceivers and the expected flow of traffic between any two nodes i and j, $\gamma_{i,j}$ are known. A fixed route between all possible source-destination pairs is also assumed to be given. This allows us to calculate the flow on the i^{th} arc, α of the network. These are standard assumptions made before addressing the capacity assignment problem for wire networks. There are however, other assumptions we will make that are particular to

the radio environment. We assume the channel is noise-free and that receivers have the ability to capture one of several signals on the channel. Our results do not depend upon the particular definition of capture and will only arise in determining the compatibility matrix discussed in the next section. We will also assume that nodes of the network are time synchronized so that all nodes share common knowledge about period beginning and ending points in the time cycle, which we will assume has a length of T time units. In general, since distances between nodes of the network vary, time synchronization implies that each slot has associated with it a guard band that is used to assure that messages lie wholly within the slot. This wasted capacity is not taken into account in our calculations.

Knowing the locations of the nodes in the network and the capture parameter of the system, allows us to generate a *compatibility matrix*. In such a matrix, a 1 in the (i, j) position indicates that arc i and arc j of the network can simultaneously transmit (specifically, their corresponding nodes being able to transmit in these directions) without causing a collision. Such an arc is said to be *enabled*. Using this compatibility matrix as the adjacency matrix of a graph, one can generate a set of *cliques* containing arcs having the property that all arcs in the same clique can be simultaneously enabled without causing a collision. If we let C_i denote the i^{th} clique, we can form a *clique cover* $C = \{C_1, C_2, \cdots, C_k\}$ with the property that every arc is contained in at least one member of C. For each clique in the clique cover one can assign an interval of time, t_i, from a given time cycle (i.e. a frame whose structure repeats), during which arcs in that clique are enabled. Since any particular arc can be contained in more than one clique, the times during which an arc is enabled depend upon the times assigned to the cliques of which it is a member.

In Figure 4.1 we exemplify this construction. In this figure, the six nodes (A through E) have arcs which are labeled 1 through 10. When node B enables arc 7, nodes A, C, and E hear the transmission and B's message is addressed to node E. During this transmission, if we assume that receivers cannot capture transmissions, one of arcs 5 or 6 can also be enabled without causing collisions. In the compatibility matrix shown in the figure, then, the 7'th row contains a 1 in the 5'th, 6'th, and 7'th positions. Each row of this compatibility matrix is generated in a like manner. Using this matrix, we have generated all possible maximal cliques $C_1, C_2, \cdots, C_{10}$, and from these have selected a particular clique cover consisting of $C = \{C_1, C_2, \cdots, C_6\}$.

If we denote the cycle time of the frame as T time units, we can allocate time durations t_i (not necessarily contiguous), where $T = \sum_{i=1}^{6} t_i$ to each of the cliques during which their arcs are enabled. Each frame then consists of a set of durations during which cliques are enabled, and the sequence of the frame cycles every T time units.



COMPATIBILITY MATRIX

1	0	0	0	0	1	0	0	0	1
0	1	0	0	1	1	0	0	1	1
0	0	1	0	0	0	0	0	1	1
0	0	0	1	0	0	0	0	0	1
1	1	0	0	1	0	1	0	1	1
1	0	0	0	0	1	0	1	1	1
0	0	0	0	1	1	1	0	0	0
0	0	0	0	0	1	0	1	0	0
1	1	1	1	1	1	0	0	1	0
1	0	0	1	1	1	0	0	0	1

$C_1 = \{1, 6, 10\}$ $C_2 = \{2, 5, 9\}$ $C_3 = \{3, 9\}$ $C_4 = \{4, 10\}$ $C_5 = \{5, 7\}$

$C_6 = \{6, 8\}$ $C_7 = \{6, 9\}$ $C_8 = \{2, 5\}$ $C_9 = \{5, 9\}$ $C_{10} = \{5, 10\}$

$$C = \{C_1, C_2, C_3, C_4, C_5, C_6\}$$

FIGURE 4.1.
Creation of the Cliques.

4.3 Queueing Approximation

In this section we will describe an algorithm for approximating the average system time (queueing plus transmission) of messages passing through a single node using spatial-TDMA. The queueing system at a given node created by this protocol consists of non-overlapping internal arrival, service, and external arrival periods which are enabled during specific periods of the frame. For example in Figure 4.2 we have shown a queue (contained in a single node of the network) with its corresponding frame, Figure 4.3. The queue contains two switches s_1 and s_2 which are used to control the internal input and service processes. At most one of these switches can be closed during any part of the frame since we do not allow simultaneous reception and transmission by radios in the network. During the first 20 time units of the frame in Figure 4.3, we see an *internal arrival interval*, during which s_1 is closed (s_2 will be open) to allow arrivals to enter from inside the network (arrivals from the attached host computer called

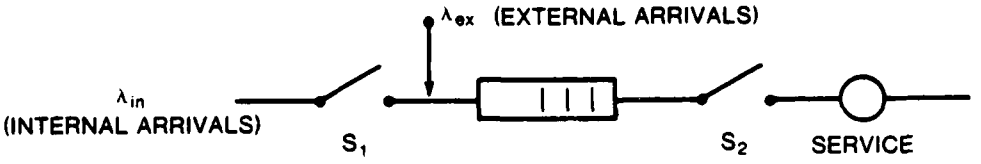


FIGURE 4.2
Model of Queue for Nodes in the Network.

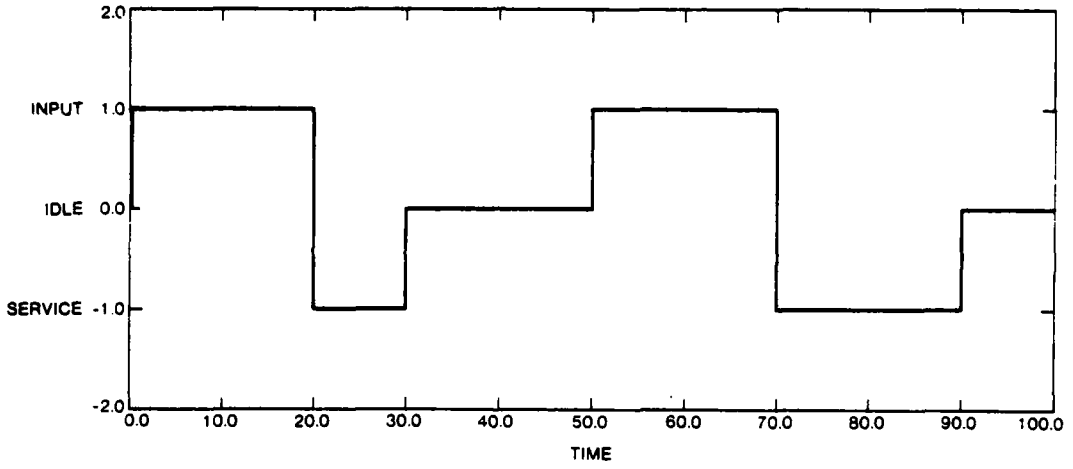


FIGURE 4.3
A Time Frame.

external arrivals can also occur during this time). These arrivals occur at rate l_{in} *message units* (assumed to be packets, bits, etc...) per time unit where we have normalized transmission rates so that each time unit can accommodate at most one message unit. We will model the internal arrival process as a truncated Poisson process. The number of internal arrivals to the queue during the first interval then is distributed as a Poisson process with parameter l_{in} subject to the

condition that no more than 20 message units arrive during this interval. In general, for an internal arrival interval of m time units, if we let $P[k\ int|\ m]$ be the probability that k internal arrivals occur in m time units, we have:

$$P[k\ int|\ m] = \frac{(m\ l_{in})^k\ e^{-(m\ l_{in})}}{k! \sum_{i=0}^{m} (m\ l_{in})^i\ e^{-(m\ l_{in})}/i!} \qquad k = 0, 1, \cdots, m$$

During the next phase of the frame, a *service interval*, of 10 units in which s_2 is closed (s_1 is open) and message units are served at the rate of one per time unit. During this time a maximum of 10 message units can be serviced. The next phase we show is an *external interval* during which both s_1 and s_2 are open and no internal arrivals or services are allowed. In the figure we have labeled the section of the frame during which both switches are open as being idle to indicate that no new messages from the network can arrive and that the node cannot transmit. This does not imply however that the are no active processes during this time. Indeed, over the entire frame as mentioned before, *external messages* (from the node's attached host) can arrive. As shown in Figure 4.2 these arrivals immediately enter the tail of the queue at a rate of l_{ex}. Again this arrival process is assumed to have Poisson statistics and thus the probability that k external arrivals occur in m time units is given by:

$$P[k\ ext|\ m] = \frac{(m\ l_{ex})^k\ e^{-(m\ l_{ex})}}{k!}$$

Switches s_1 and s_2 are then turned on and off according to the time patterns depicted in Figure 4.3 and continue to cycle every 100 time units. Observe that there is no limit on number of messages that can arrive from the attached host computer over the course of a frame as there is from messages arriving from inside the network.

An exact analysis for the average system time has not been carried out as of yet and thus we seek an approximate formula. Fortunately a fluid approximation [Klei76a] to this system gives very good results and we will illustrate this method using Figure 4.4. In the fluid approximation, waiting times are calculated by assuming the actual backlog of packets in the queue is approximated by the average backlog.

In figure 4.4 we have plotted the growth of the average backlog of message units in the system during the course of the frame shown in figure 4.3. For reasons to be explained later, we have started this growth pattern at the beginning of the last external interval (i.e. at t=90). During this interval, since internal arrivals are prohibited, only external arrivals can add to the backlog. If we assume the rate at which the queue increases during an external period to be M_{ex}, we see that the expected backlog grows linearly with this slope during this interval. As the frame progresses to the first interval, the growth rate for the backlog is given by M_{in}, which is the sum of M_{ex} and the additional rate offered by the internal arrivals. A service period follows in the next interval and the backlog drops by a rate of $M_s < 0$ message units per time unit. Since external messages can arrive during any type of interval, M_s is equal to M_{ex} less

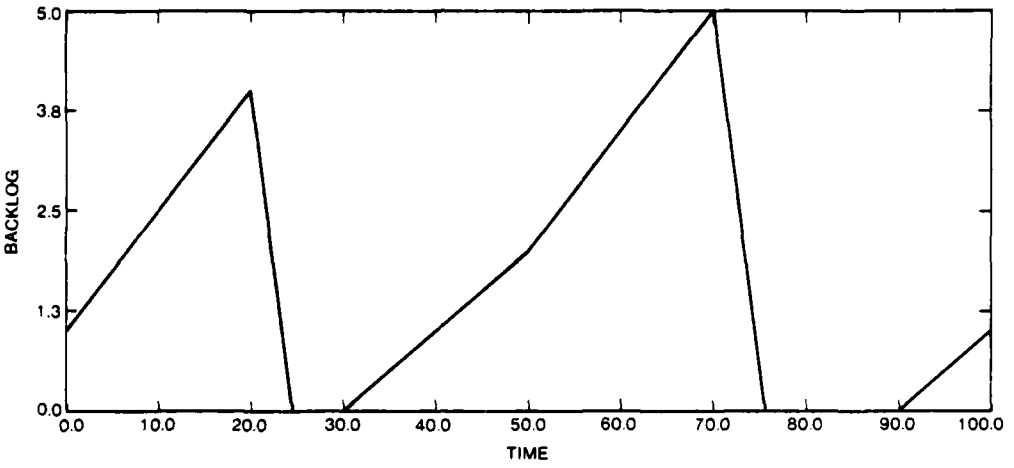


FIGURE 4.4
The Backlog for Figure 4.3.

the service rate of 1 message unit per unit time. As seen in the figure, the backlog of message units in the queue drops to zero before the service interval is finished. Internal arrivals during this time are assumed to be immediately serviced and thus do not contribute to the backlog. This process continues in this manner until the last external period, at which point it starts from a zero backlog once again. We will call a point on the frame a *zero-point* (it is a regenerative point for the frame) if, starting with an empty queue at this point, yields a backlog of zero after exactly one frame.

We now calculate the rates described in the previous section. First we make the following definitions:

Let T = Length of the time frame
T_{ex} = Total time of external intervals per frame
T_{in} = Total time of internal arrival intervals per frame
T_s = Total time of service intervals per frame

From our previous discussion we can write:

$$l_{ex} = L_{ex}/T$$
$$l_{in} = L_{in}/T_{in}$$
$$M_{ex} = l_{ex}$$
$$M_{in} = l_{in} + M_{ex}$$
$$M_s = M_{ex} - 1$$

The area under the backlog curve represents the number of message-time units accumulated by messages arriving during the frame. Dividing this by the average number of message units that arrive during the frame, $L_{in} + L_{ex}$, gives, by Little's result, [Litt61a] the average time spent in the queueing system. Because we began the calculation at a zero-point, message delays for all arrivals to the frame are counted. It is clear we can always find a zero-point on any frame satisfying $T_s \geq L_{in} + L_{ex}$.

This then describes the algorithm for the approximation which we summarize as:

1. Find a zero-point.

2. Calculate the area under the backlog curve.

3. Divide by $L_{in} + L_{ex}$ to arrive at the average system time.

It is clear that the ordering of the intervals and their lengths greatly influences the average system time. Suppose, for example, that we change the frame in Figure 4.3 by coalescing all the service and intervals together and placing them on the frame as shown in Figure 4.5. Although it has exactly the same interval lengths, this translation increases the average system time (in fact it is a worst case example). In this system, any messages that arrive during the internal arrival interval must wait at least 30 time units before being serviced. If the service and internal arrival intervals are interchanged, it is clear that the average system time would decrease by at least this much. In fact, the system that has the minimum average system time is one that spreads arrival and service intervals in infinitesimal units that alternate with each other. In this way, an arrival is immediately serviced in the following interval after accruing little waiting time. In a practical implementation of spatial-TDMA, however, there are limits to how small one can make interval sizes since radios have a finite switching time between transmission and reception, and choosing a frame pattern that minimizes the system time for one particular queue in the network does not in general decrease the total average message delay for messages in the entire network. In fact, finding the alternation that does achieve the minimum delay for all message in the network is a very difficult problem which we will address in a later section.

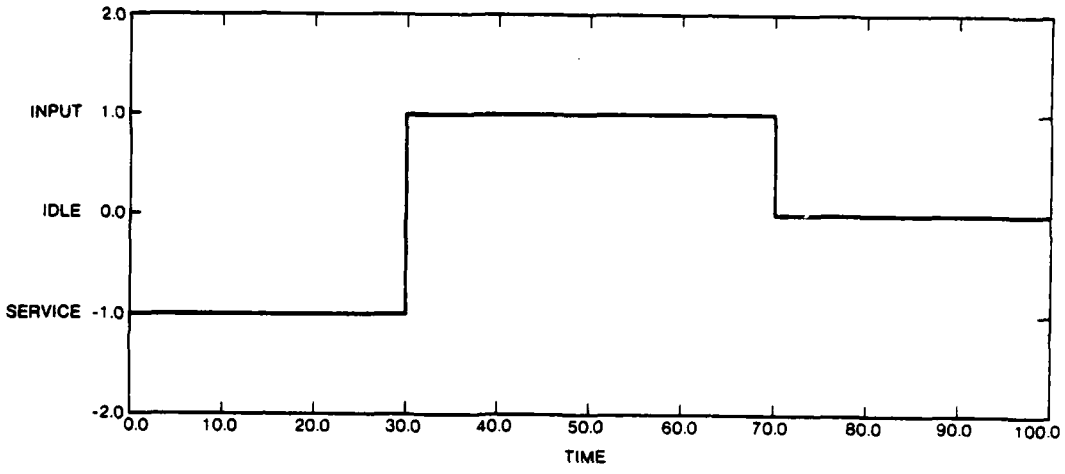


FIGURE 4.5
A Worse Case Cycle.

4.4 Discussion of Delay Results

In this section we compare simulation results with those found using the fluid approximation described in the previous section. Numerous frames were randomly generated that had the same input parameters (T, T_s, T_{in}, T_{ex}, L_{ex}, and L_{in}) and results of the simulation checked against those of the approximation. Three such frames are shown in Figure 4.6 where +1 steps correspond to internal arrival intervals, 0 to external intervals, and −1 to service intervals. In all three frames T=10000, T_{ex}=6000, T_{in}=2000, T_s = 2000, the average service and internal arrival intervals have length 200, and the average external arrival rate is 200 message units over the frame time T. The mean service time as a function of ρ for these frames is shown in Figure 4.7.

There are several interesting features of these curves. We first see the close match between the simulation and approximation thus assuring us that the approximation is accurate. The variation between the mean system time for the three frames is very large which shows the dependency upon the ordering and size of the intervals of the frame. For example, for ρ=.7, set 1 has a mean system delay of about 650 time units whereas set 3 has a value of 2800, more than four times as much. The extreme delays of the third frame arise from the long periods (9000, 10000) and (0, 2000), during which there are no service intervals. All packets arriving during these periods create a backlog that cannot be depleted until much later in the frame. On

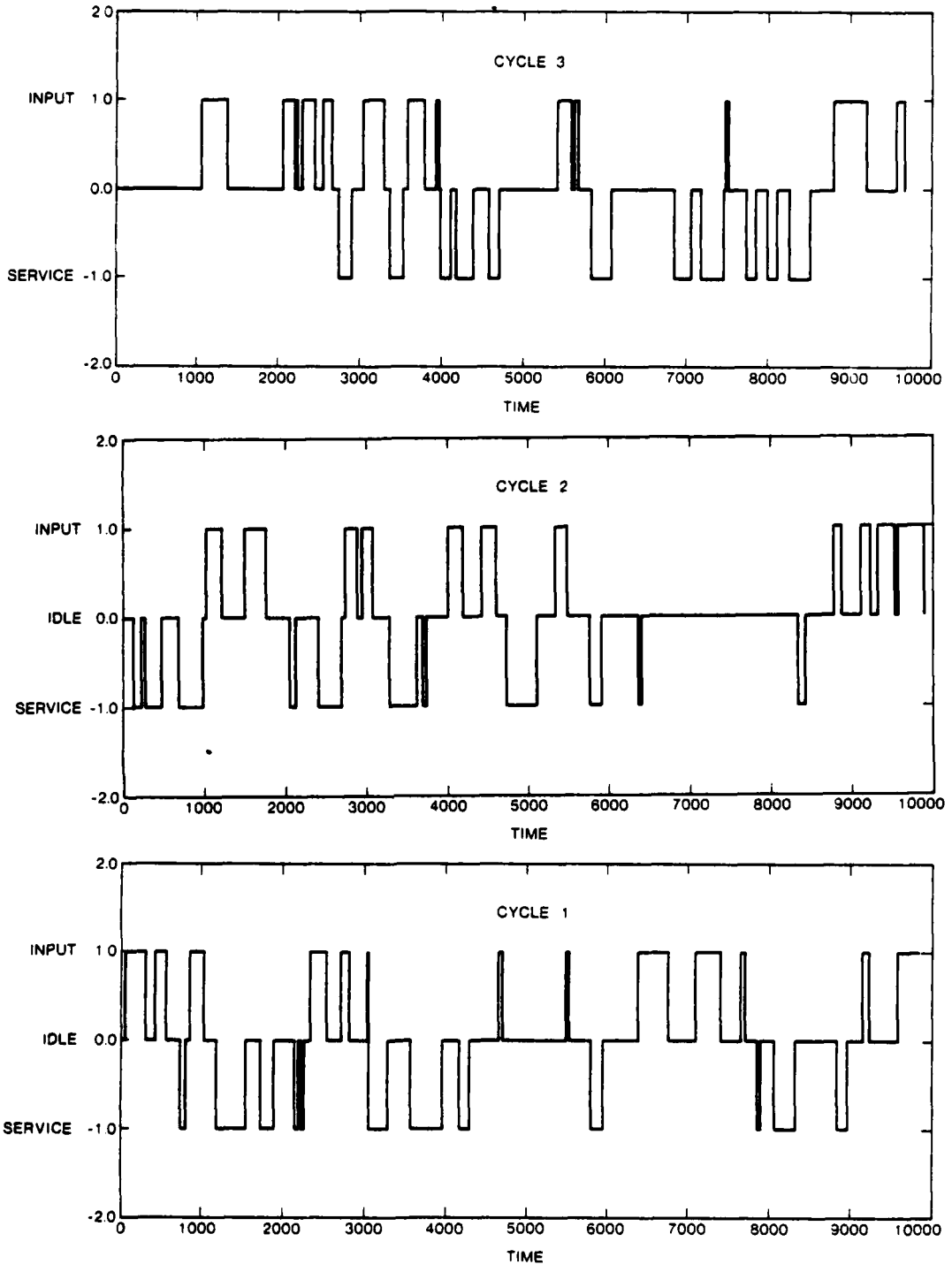


FIGURE 4.6
Three Randomly Generated Frames.

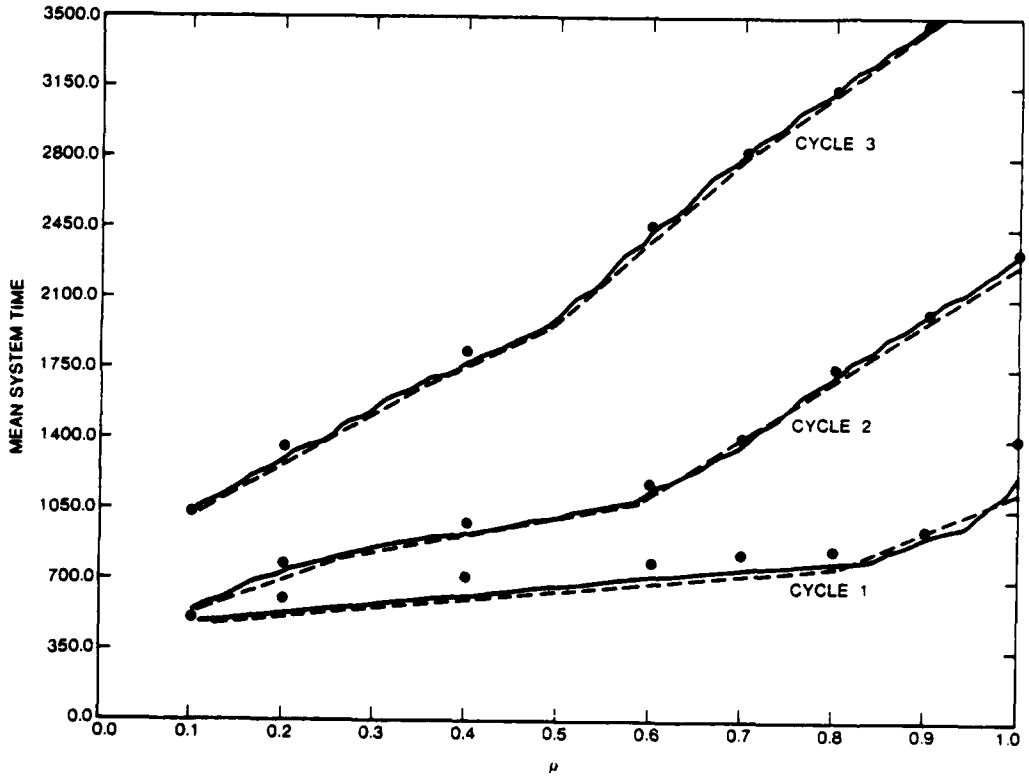


FIGURE 4.7
Curves for the Three Frames of Figure 4.6.

the other hand, the fortuitous placement of intervals in frame 1 consists of groups of arrival intervals followed by service intervals that allow an accumulated backlog to be serviced quickly.

Even though there is a large variance in the curves, there is a striking similarity in the shapes of the curves. The curves are very well approximated by a piece-wise linear function (shown as a dashed line in the figure). We note that they do not approach infinity at $\rho = 1$. Since other queueing systems have explosive growth as $\rho \rightarrow 1$, arising from the omni-present $1/(1-\rho)$ term in most queueing equations, this behavior is very unusual. The explosive growth for most queueing systems arises, however, from the fact that there is variation in the arrival pattern of messages to the queue, and as $\rho \rightarrow 1$ the probability that a sequence of arrivals saturates the queue approaches 1. In spatial-TDMA however, the number of arrivals for each internal arrival interval is constrained to be no greater than the normalized interval length, and thus the variance of arrival statistics is also constrained. Besides reducing queueing delays, this restriction on the number of messages that can enter the system during an interval also acts as

a natural flow control mechanism for messages in such a network.

The slope changes in the piece-wise linear approximation occur when the arrival rate is so large that a sufficient number of arrivals to an internal arrival interval cannot be serviced in the next set of service epochs. For example, the change about the point $\rho = .6$ for frame 2 arises from the fact that for $\rho > .6$, some arrivals over the interval (6000,7800) must wait until the next service set of intervals (1200, 2300) to be processed. For lesser values of ρ, $\rho < .6$, most arrivals to (6000, 7800) are serviced in the interval (7800, 9100) and thus suffer less delay. Naturally as ρ increases, the proportion of messages that must wait until (1200, 2300) to be serviced grows and so does the mean system time. Each of the breaks in the piece-wise linear approximation can be explained in this manner.

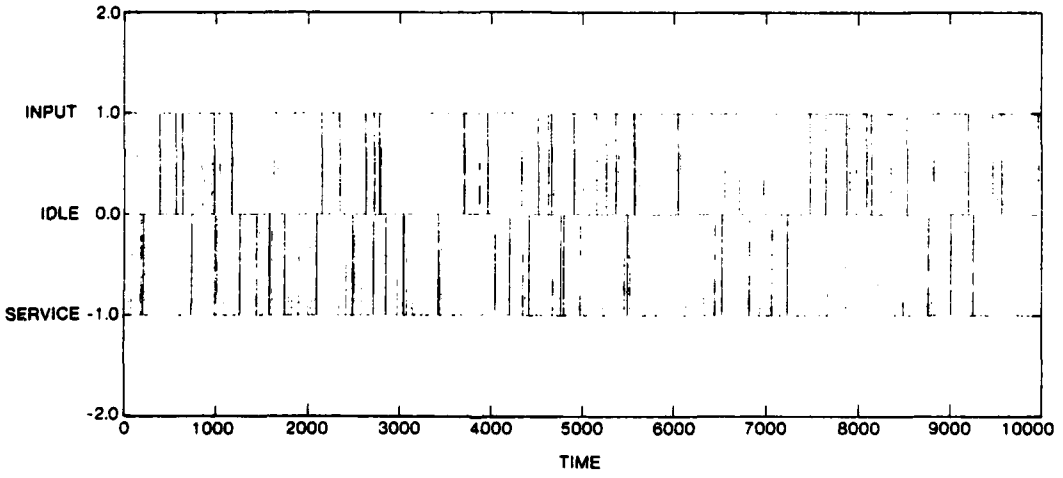


FIGURE 4.8
Reducing the Average Period Size.

In Figure 4.8 we have shown a frame for the same input parameters but where the average size of the service and internal arrival intervals is equal to 20 time units instead of 200 as in those of Figure 4.6. The corresponding mean system time curve is shown in Figure 4.9. We see a marked decrease in the mean system time for this frame in comparison to the previous set of frames. This demonstrates the dependency of the mean system time upon the size of the intervals. If we adjust the frame to minimize the mean system time, as shown in Figure 4.10 (where for illustrative clarity we have only shown a portion of the frame), the resultant delay is approximately equal to 1 time unit throughout the entire range of ρ. For such a frame the majority of the arrivals to the system are immediately serviced in the following service interval.

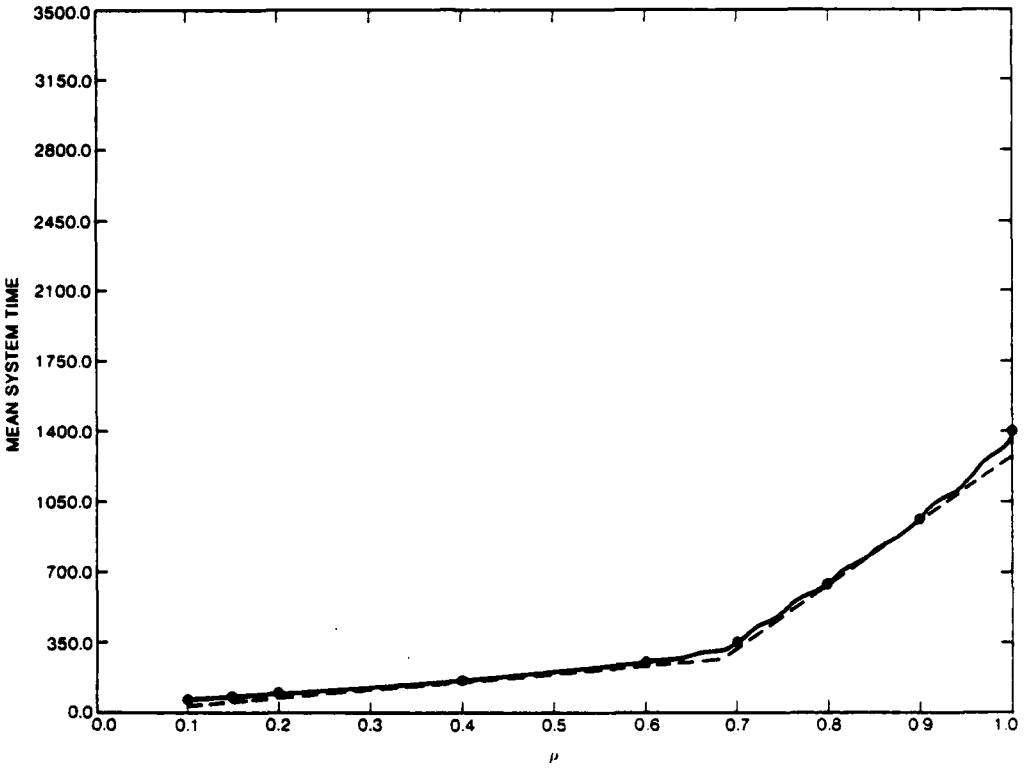


FIGURE 4.9
The Curve for Figure 4.8.

4.5 The Capacity Assignment Problem

4.5.1 Introduction

In the previous sections we determined, for a given frame, a good approximation for the average delay messages experience in passing through a network using spatial-TDMA. We observed that the ordering, size, and number of periods from the frame allocated to internal and external arrivals and service periods, had a great influence on the mean delay. One observation we made was that, in general, the mean delay decreased as the average size of the periods decreased. In practical implementations of this protocol there is a minimum size for any period which is determined by the switching time for radio is in the network and by overhead considerations. Let us therefore suppose that all periods in a frame have a length

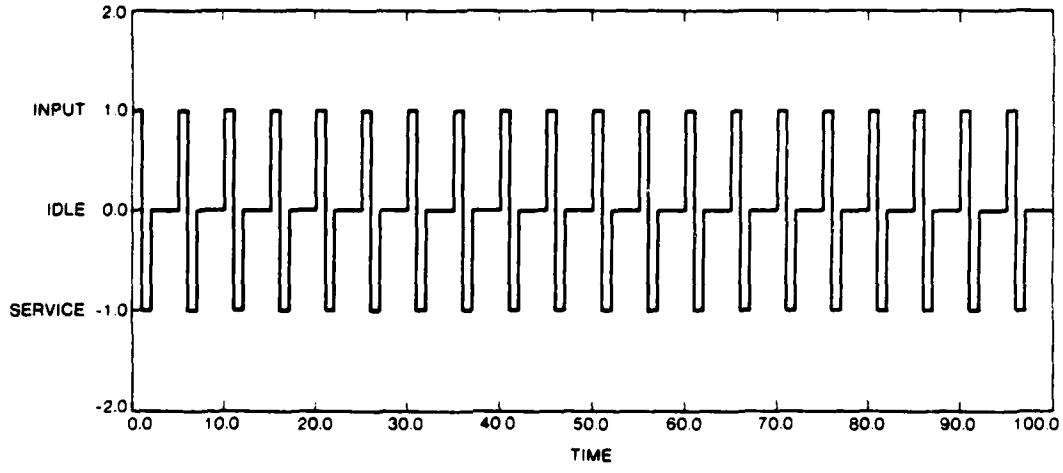


FIGURE 4.10
An Optimal Frame.

determined by this minimal switching time, and let us normalize the frame duration, T, to this slot size. We will say that one message unit can be transmitted during this duration. With this restriction, then, every frame consists of a number of unit slots assigned to the cliques of the clique cover in some particular permutation. This restriction does not effect the optimality of frame selection to any great degree, since the unit slot is, in the continuous frame case, a lower bound to slot size anyway, and is usually very small in relationship to T. We will therefore, in a later section, take the liberty to refer to the size of the intervals T_{ex}, T_{in}, and T_s as being continuous parameters when we pose a convex programming problem.

Before beginning to develop the capacity assignment problem, let us first establish our notation. Suppose we are given a clique cover, $C = \{C_1, C_2, \cdots, C_k\}$, and a time vector, $\underline{t} = (t_1, t_2, \cdots, t_k)$, where t_i is the time from the frame (having duration T) allocated to C_i. We will denote $\| \underline{t} \| = \sum_{i=1}^{k} t_i$ and will say a time vector $\underline{t}$ is feasible if $\underline{t} > 0$, $\| \underline{t} \| \leq T$, and certain flow constraints, which are defined later, are satisfied. Let the total flow of messages into the network per frame be given by γ and let α_i be the flow of messages into queue i over that time. Let q be the number of queues in the network. For a given feasible time vector, $\underline{t}$, and queue i, let $D_i(\underline{t})$ be the average delay of messages passing through queue i, and let T'_{in}, T'_{ex}, and T'_s be the total amount of time from the cycle allocated for the internal, external, and service periods respectively. Also for queue i let L'_{in} and L'_{ex} be the average traffic flow from internal and external sources over a time frame. It is clear that $\alpha_i = (L'_{in} + L'_{ex})/T$ and we will define the $(k \times 1)$ binary vectors $\underline{M}'_{in}$, $\underline{M}'_{ex}$, and $\underline{M}'_s$ to satisfy $T'_{in} = \underline{M}'_{in} \cdot \underline{t}$, $T'_{ex} = \underline{M}'_{ex} \cdot \underline{t}$, and $T'_s = \underline{M}'_s \cdot \underline{t}$. Observe that we can form a matrix M' from these vectors that will satisfy $M' \cdot \underline{t} = (T'_{ex}, T'_{in}, T'_s)$.

With these definitions we can define the capacity assignment problem formally as:

$$\underset{\underline{t},\ \text{All Possible Frames}}{\text{Minimize}} \sum_{i=1}^{q} \frac{\alpha_i}{\gamma} D_i(\underline{t})$$

$$\begin{aligned} \text{subject to } & \underline{t} \geqslant 0 \\ & ||\underline{t}|| - T \leqslant 0 \\ & \underline{M}_{in}^{i} \cdot \underline{t} - L_{in}^{i} \geqslant 0 \quad i = 1, 2, \cdots, q \\ & \underline{M}_{s}^{i} \cdot \underline{t} - L_{in}^{i} - L_{ex}^{i} \geqslant 0 \quad i = 1, 2, \cdots, q \end{aligned}$$

Before discussing the complexity (and mathematical intractability) of solving this problem, let us first establish that the feasibility region is convex. Suppose that $\underline{t}$ and $\underline{t}'$ are two feasible time vectors. To show convexity we need to show that $p\underline{t} + (1-p)\underline{t}'$ where $0 \leqslant p \leqslant 1$ is also feasible. This is obvious however, since, using the first constraint:

$$\underline{M}_{in}^{i}(p\underline{t} + (1-p)\underline{t}') = p\underline{M}_{in}^{i} \cdot \underline{t} + (1-p)\underline{M}_{in}^{i} \cdot \underline{t}' \geqslant L_{in}^{i}$$

The same argument holds for the second constraint.

The complexity of solving this capacity assignment problem arises from the fact that the optimization occurs over all feasible time vectors, $\underline{t}$, *and* all possible ways of allocating periods in frames satisfying $\underline{t}$. Even if we were given (by some divine mathematician!) the time vector which was optimal, finding the frame (the ordering of all T slots) that achieves minimal delay is in itself mathematically intractable. In the following sections we will present an approach that finds optimal time vectors over the class of frames that have been randomly selected, but first we will address the question of feasibility.

4.5.2 Feasibility

In this section we establish a procedure for determining if a given clique cover, for a given set of queues and corresponding flows α_i, permits a feasible time vector. It is easy to create cases where, no matter how one adjusts the components of a time vector, some queues will have more flow into them than they can accommodate. We have already seen that the

feasible region is convex. This allows us then to formulate the feasibility problem as a linear program:

$$\underset{\underline{t}}{\text{Minimize}} \; ||\underline{t}||$$

$$\begin{aligned} \text{subject to } & \underline{t} \geq 0 \\ & \underline{M}'_{in} \cdot \underline{t} - L'_{in} \geq 0 \quad i = 1, 2, \cdots, q \\ & \underline{M}'_{s} \cdot \underline{t} - L'_{in} - L'_{ex} \geq 0 \quad i = 1, 2, \cdots, q \end{aligned}$$

If the solution to this linear program has $||\underline{t}|| < T$ than any vector with $\underline{t} + \underline{\epsilon}$ where $\underline{\epsilon} > 0$ and $||\underline{\epsilon}|| = T - ||\underline{t}||$ is a feasible time vector, otherwise no feasible time vector exists. We might add that phase 1 of solving this linear program can easily be solved by a time vector having all components equal to $t_i = \underset{i}{\text{Max}}\,(L'_{in} + L'_{ex})$, and that phase 2 of the program can halt as soon as the objective function $||\underline{t}||$ becomes less than T. Thus this feasibility program can be efficiently solved.

4.5.3 Average of All Randomly Generated Frames

Due to the mathematical intractabilities of the frame assignment problem, we seek in this section to create an approximation to the average delay of messages in passing through a queue in which the alterations of frame periods have been randomly selected. This will represent an upper bound to the mean delay obtained with an optimal ordering and, as we will see, will lead to a tractable program. Let us therefore intuitively motivate our approximation to this delay. Recall that we have normalized the time axis into time slots equal to the transmission of one message unit. Suppose we are given T_{ex}, T_{in}, and T_s and have selected one of the possible random frames having these time values. Select a point in this frame, T' where $T' < T$ and let T'_{ex}, T'_{in}, and T'_s be the number of external, internal, and service slots used in generating the frame up to time T'. Since the frame was randomly generated, the next slot, $T'+1$, will be drawn with a uniform probability from the remaining population of slots. For the purpose of clarity, we will focus our discussion in terms of service slots (analogous statements can be made for external and internal arrival slots). At the $T'+1$ step, the probability that the next slot selected is a service period is given by $P'_s = (T_s - T'_s)/(T - T'_s - T'_{ex} - T'_{in})$. The original probability of selecting a service interval was given by $P_s = T_s/T$, and thus we see this probability changes as the cycle is generated. If T is large however, over most of the frame, one would expect that the probability of selecting a service period at each step does not differ much from P_s. Intuitively, T_s, T_{in}, and T_{ex} decrease at a rate in proportion to their sizes, and since their sizes are assumed to be large, their relative proportions do not alter significantly over most of the frame. Only when most of the slots have been given out, when T' is close to T, will we expect a wide variation in P'_s. The main point is that until this time, P'_s has remained close to

P_s and the major component of the delay, as determined by the backlog curve of the fluid approximation, has already been determined. We thus conclude that the delay obtained from generating frames using the initial probabilities, P_s, P_{in}, and P_{ex} throughout the frame generation will not differ much from those generated by allowing P_s to change throughout frame generation.

Now that we have motivated the approximation, let us describe it more precisely. Suppose we are at a height k on the backlog curve. From the previous discussion, the probability that the next slot is an external, internal, or service slot is given by $P_{ex}=T_{ex}/T$, $P_{in}=T_{in}/T$, and $P_S=T_s/T$ respectively. Recall that $l_{in}=L_{in}/T$ is the probability that a randomly selected slot contains an internal arrival. Define u_i $i=1,2$ to be the probability that i arrivals come into the system in the next slot, and let d be the probability that a message leaves the system in the next slot. Using the above definitions we can then write:

$$\begin{aligned} u_1 &= P_{ex}\, l_{ex} + P_{in}\,[\, l_{ex}(1-l_{in}) + l_{in}(1-l_{ex})\,] \\ u_2 &= P_{in}\, l_{in}\, l_{ex} \\ d &= P_s(1-l_{ex}) \end{aligned}$$

and the evolution of the height of the backlog curve is given by a random walk with the above probabilities. The state transition graph for this random walk is given in figure 4.11.

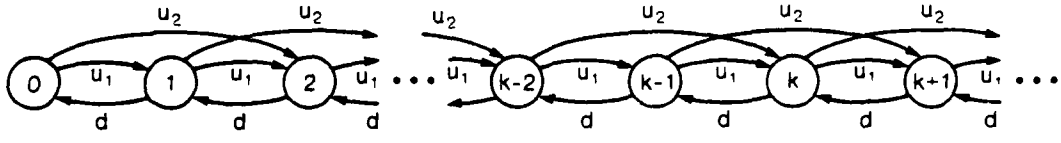


FIGURE 4.11
State Transitions For the Random Walk.

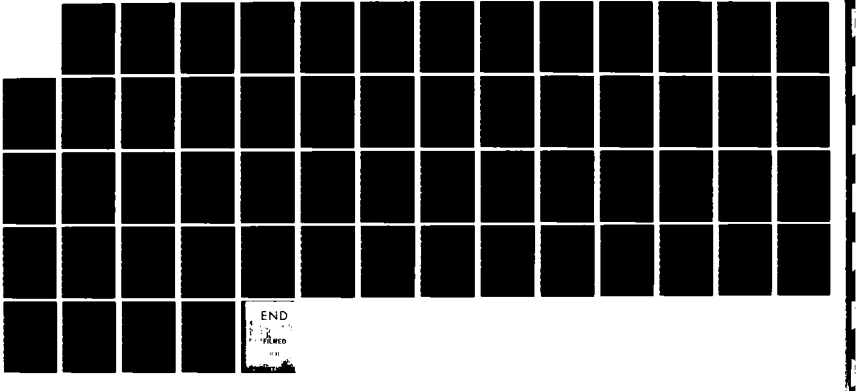
The state transition equations are given by:

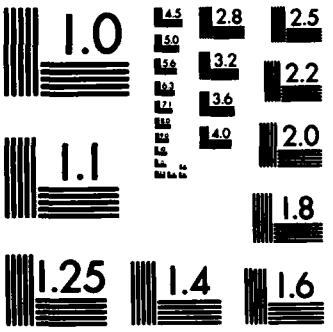
$$\begin{aligned} (u_1+u_2)P_0 &= dP_1 \\ (u_1+u_2+d)P_1 &= u_1P_0 + dP_2 \\ (u_1+u_2+d)P_k &= u_2P_{k-2} + u_1P_{k-1} + dP_{k+1} \quad k\geq 2 \end{aligned} \tag{4.1}$$

where P_k is the probability that the system is in state k. Before solving these equations, let us determine how we will use the solution to generate our approximation to the average delay. We will proceed in the same way as in the fluid approximation, namely finding the area under the backlog curve and then dividing by the average number of customers that entered the system. Suppose we knew the steady state probabilities P_k of being at step k in our approximation. Points at a height k that move to $k+1$ incur an increase in the area that is equal to $k+.5$ where the k arises from the rectangle of height k and width 1, and the .5, from the

AD-A133 554 ADVANCED TELEPROCESSING SYSTEMS(U) CALIFORNIA UNIV LOS ANGELES DEPT OF COMPUTER SCIENCE L KLEINROCK 30 SEP 82 UCLA-ENG-83-24 MDA903-82-C-0064 2/2

UNCLASSIFIED F/G 17/2.1 NL





triangle having unit width and height. Similar calculations can be made for other steps and we can write the average area for each step as:

$$A = P_0(u_2+.5u_1) + \sum_{i=1}^{\infty} P_i[u_1(i+.5) + u_2(i+1) + d(i-.5)] \tag{4.2}$$

$$= P_0(u_2+.5u_1) + \bar{P}(u_1+u_2+d) + (1-P_0)(u_2+.5u_1-.5d)$$

In these equations $\bar{P}$ is the average value of k. We can calculate both $\bar{P}$ and P_0 by using standard transform methods on equations (4.1) [Klei76a]. Defining the z-transform of P_i to be $P(z) = \sum_{i=0}^{\infty} P_i z^i$, we have:

$$P(z) = \frac{2u_2+u_1-d}{u_2z^2+(u_1+u_2)z-d}$$

from which, using standard theory we can derive:

$$\bar{P} = \frac{3u_2+u_1}{d-2u_2-u_1}$$

$$P_0 = \frac{d-2u_2-u_1}{d}$$

and using these in equation (4.2) yields:

$$A = \bar{P}(u_1+u_2+d) \tag{4.3}$$

Thus we can calculate the average delay by dividing by the average number of messages that arrive in a randomly selected slot to determine the following delay equation:

$$D(T_{ex}, T_{in}, T_s) = \frac{T}{(L_{ex}+L_{in})} \frac{(3u_2+u_1)}{(d-2u_2-u_1)} (u_1+u_2+d) \tag{4.4}$$

We have denoted $D(\cdot)$ as being a function of three variables for the purpose of clarity. Since $T = T_{in} + T_{ex} + T_s$ only two of the variables are independent.

4.5.5 Comparing the Approximation with Randomly Generated Frames

In this section we want to show the results when equation (4.4) is compared with data obtained when we generate random frames and calculate their mean delay using the fluid approximation. Our procedure was, for a given set of parameter values L_{ex}, L_{in}, T_{ex}, T_{in}, and T_s, to generate 1000 random frames and determine the minimum, maximum, mean, and variance of the delays for these samples. We then compared the average delay to equation (4.4) to determine if the approximation was close. We performed this procedure for many different

selections of parameter values and all showed similar behavior. To demonstrate this behavior we have selected two sets to plot. Each of these sets had T=1000. Set 1 has T_{in}=400, T_{ex}=100, T_s=500, L_{ex}=100, and we varied L_{in} over the range $20 \leqslant L_{in} \leqslant 380$. Set 2 had T_{in}=300, T_{ex}=200, T_s=500, L_{in}=200 and L_{ex} varied over $20 \leqslant L_{ex} \leqslant 280$. In figure 4.12 we show how the approximation fared in relationship to the mean of the generated frames.

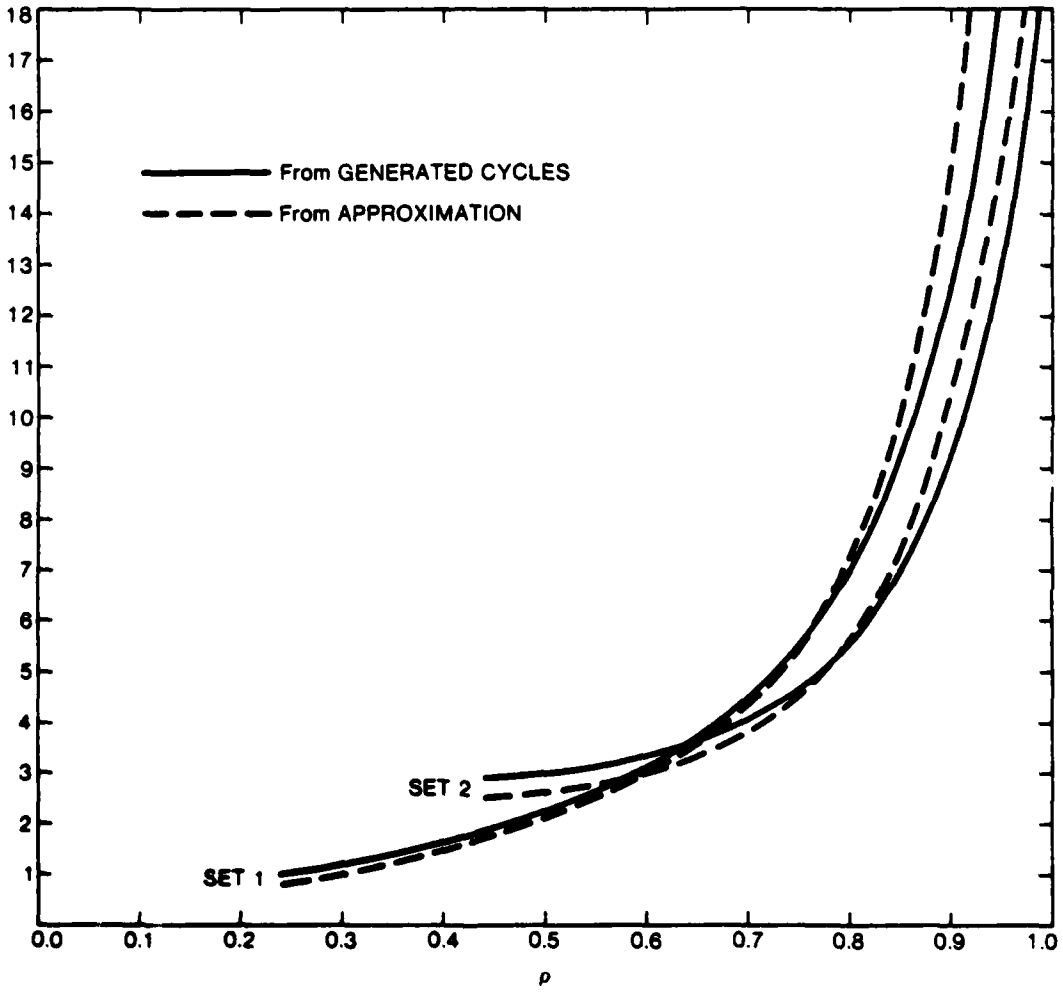


FIGURE 4.12
Validating the Approximation.

We see in this figure that the approximation is very close to that of the sample mean and that only for very large values of ρ does it break away from the generated frames. This is explained by the fact that the approximation is "more stochastic" than the randomly generated frames

since it is given by a random walk, and that for high ρ values the random walk tends to spend a lot of time "attempting to escape to infinity" in contrast to the generated frames that have an upper bound to their maximum height.

We can extract more information about the randomly generated frames by looking at Table 4.1.

ρ	Mean	Variance	Min Delay	Max Delay
0.240000	1.019933	0.059683	0.882035	1.292519
0.280000	1.173994	0.076922	1.000477	1.418889
0.320000	1.345148	0.097195	1.122492	1.737658
0.360000	1.519549	0.123856	1.228147	2.035303
0.400000	1.723441	0.145813	1.376908	2.425304
0.440000	1.919901	0.193245	1.435101	2.563087
0.480000	2.158198	0.218624	1.594046	3.138185
0.520000	2.434490	0.241395	1.902604	3.434316
0.560000	2.769812	0.325750	1.993773	4.291554
0.600000	3.172192	0.414819	2.390554	4.736672
0.640000	3.621928	0.522270	2.618372	5.937060
0.680000	4.222784	0.670217	2.910517	7.552518
0.720000	4.881158	0.869889	3.210058	9.927224
0.760000	5.741120	1.051270	3.566030	9.822580
0.800000	7.005569	1.534562	3.751551	15.805325
0.840000	8.766911	2.257944	5.079778	19.209290
0.880000	11.501342	3.460731	6.268044	32.473289
0.920000	15.094280	4.126865	7.613554	34.197002
0.960000	22.544538	7.200002	9.852835	65.090919

TABLE 4.1.

In this table we list, for given values of ρ, for set 1, the mean delay, variance, minimum delay, and maximum delay that was found over the generated frames. We see that although the difference between the minimum and maximum delays are often quite substantial (especially for large ρ) the variance is usually very small. This implies that the mean of all the generated frames is not much different from the mean of a particular generated frame. Another way to see this is to plot the coefficient of variation (defined to be the variance divided by the mean) as a function of ρ. In figure 4.13 we plot this, as well as the variance, and see that the coefficient of variation is quite small throughout all ranges of ρ. Using this approximation we are now in a position to formulate the capacity assignment problem for networks of this type.

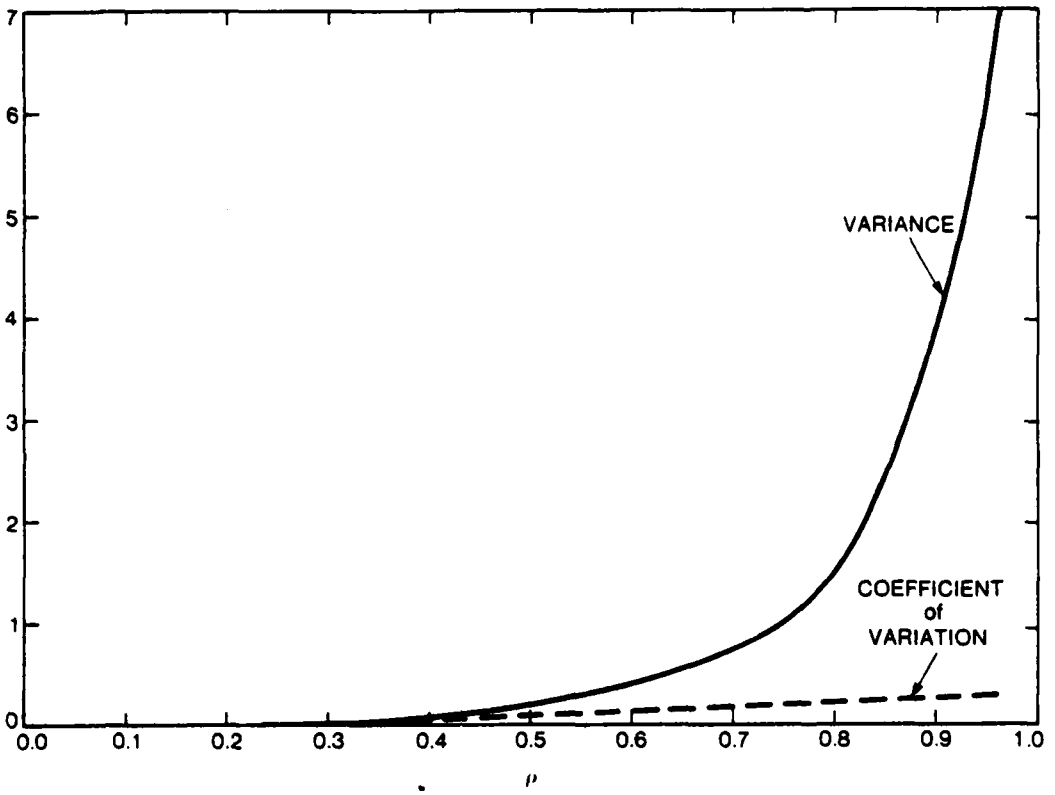


FIGURE 4.13
The Coefficient of Variation as a Function of ρ.

4.5.5 The Capacity Assignment Problem for Random Frames

In this section we will create the capacity assignment problem for spatial-TDMA networks in which the time slots to cliques have been performed in a random manner. We first

suppose that there exists a feasible time vector. Knowing this we can then formulate the capacity assignment problem as:

$$\underset{\underline{t}}{\text{Minimize}} \sum_{i=1}^{q} \frac{\alpha_i}{\gamma} D_i(M^i \cdot \underline{t})$$

$$\begin{aligned} \text{subject to } & \underline{t} \geqslant 0 \\ & ||\underline{t}|| - T \leqslant 0 \\ & \underline{M}_{in}^i \cdot \underline{t} - L_{in}^i \geqslant 0 \quad i = 1, 2, \cdots, q \\ & \underline{M}_s^i \cdot \underline{t} - L_{in}^i - L_{ex}^i \geqslant 0 \quad i = 1, 2, \cdots, q \end{aligned}$$

Once again the feasibility region is convex and it can be also shown, by differentiating $D_i(\cdot)$, that the Hessian matrix of second partials is positive semi-definite and thus the objective function is also convex. We thus have a convex programming problem which can be solved using any number of well-known solution techniques [Avri76a].

4.6 Conclusions

The flexibility offered by broadcast radio presents clear advantages over the use, in conventional networks, of point to point cable. Besides being more adaptable to changing topologies (and as demonstrated in this paper, changing traffic characteristics) radio does not require the construction of elaborate supporting mechanisms for their implementation (this is not to imply that packet radio networks do not have their technical problems!). The design of networks that use spatial-TDMA is facilitated since precise specifications of topologies and traffic characteristics do not initially need to be known. Very often the true flow of traffic between nodes of a network are only known after the network has been used for a period of time. As shown before, incorrect estimations of the traffic flows can be disastrous for a wire network. With radio however, network designers have the ability to simply alter capacities to match the actual traffic flows. A straightforward approach based on the results of this paper, would be to start with an initial guess of the traffic flows. These would be used to determine the initial capacities, and the system would be allowed to operate for a period of time during which statistics of the actual traffic flows would be gathered. These more precise estimations would then be used to again assign capacities. This process could be refined further, where perhaps several capacity assignment *plans* could be incorporated to match the flows during different parts of the day, of the week, of the year etc... since traffic patterns frequently change according to a predictable schedule (for example there is very little traffic flowing from the east coast after 5 p.m.). Scheduling and synchronizing changes between different plans, would be a non-trivial implementation problem.

The above advantages of radio based networks can also be combined with the results of the extensive research into the behavior of wire networks. Since with spatial-TDMA, radio networks simulate wire networks, many of these results can be readily adapted to the radio environment. For instance, routing and flow control techniques designed for wire networks can

be run without modifications on the radio net. In particular, the flow-deviation algorithm [Frat73a] could be used to calculate an optimal (minimal average delay) routing policy. In this procedure, flows on the arcs are altered (preserving the flow required between source-destination nodes of the net) in a fashion that lowers the average delay until a minimum is found. Since the function for the average delay is different in our case than that of the standard flow-deviation algorithm, we would have to modify the algorithm slightly but the overall strategy would be similar since the equation for the mean delay is convex. After executing this algorithm, the flows on the arcs would be changed, and thus we could run the capacity assignment algorithm specified in this paper to re-assign capacities to the arcs. After this new capacity was assigned, it would be followed by the flow-deviation algorithm again and this alteration would continue until the change in the average delay became smaller than some specified value. This *Capacity and Flow Assignment* problem has been developed for wire networks [Klei76a] and one could adapt similar procedures to be used in the radio environment.

Thus it would appear that radio networks using spatial-TDMA have both the advantages of a flexible topology and that of being able to draw from the wealth of results created for conventional wire networks. It also however suffers from some of the disadvantages of wire based networks. In both spatial-TDMA and wire networks, nodes are assumed to be stationary and besides help from adaptive routing procedures, there is no allowance for dynamic changes to the capacities of the arcs of the network. Indeed it would be somewhat tragic if the most salient feature of the radio media, its ability to adapt to topological changes, was not taken into account in the design of the channel access scheme of a packet radio network. The channel access scheme for such a network would need to account for changing topologies and traffic demands. Certainly the access scheme as presented in this paper does not have these characteristics. However by modifying the scheme in a manner which allows dynamic slot adaptations, in a local fashion, to topological and traffic demands, spatial-TDMA can be made to operate in such a mobile environment.

Another drawback of spatial-TDMA concerns the nature of the traffic flowing in the net. Since arcs are enabled for specific periods of time during the cycle, efficient use of the channel requires that messages be available for transmission during these times. To elucidate the traffic characteristics most suitable for spatial-TDMA, imagine that traffic arrives to the arcs in the network in a steady flow. The queue sizes at each arc then would never grow large because the queue could be steadily decreased during the transmit periods of the cycle for that arc, and we would expect the message delay to be small. In contrast to this, if we imagine the same average flow arriving to the arcs in large bulks, with long interarrival idle periods, large queue sizes are formed which, because the arc only has fixed time from the cycle in which to transmit, would take a long time to empty. This would then result in large message delays. We can thus conclude that spatial-TDMA is most suited for steady streams of traffic. We might note here that random channel access protocols are most suited for the case of bursty traffic. As an example of a network where spatial-TDMA would be suitable, suppose that terminals were connected to nodes of a spatial-TDMA network. If there was a sufficient number of

terminals connected to each node, the net total flow from all these bursty sources, according to the law of large numbers, would be steady and thus suitable for our proposed protocol. In fact, any method which concentrates many bursty sources into one traffic stream will allow spatial-TDMA to be effectively utilized.

CHAPTER 5
RUDE–CSMA

So far we have considered the interference problem under the conditions of using minimal information (in Chapter 2) and of using global information (Chapters 3 and 4). In this section we consider an intermediate approach to this problem that uses local information which nodes obtain when they sense the channel. This investigation leads us to define a two-parameter family of protocols for the multi-hop environment called rude-CSMA.

5.1 Introduction

Let us first review the CSMA protocol as defined for the single-hop environment that was discussed in Chapter 1. It is well known that CSMA [Klei75a] is an efficient channel access protocol for single-hop packet radio environments. In this protocol, nodes of the network sense the channel prior to transmitting packets. If the channel is sensed busy, the sensing node refrains from transmitting (to avoid a collision) and re-schedules its transmission until a future time. In a multi-hop environment sensing the channel provides information only about a subset of the nodes in the network and thus a transmission by one node that is surrounded by silent nodes is not guaranteed to be successful. An example of this is shown in figure 5.1. In this figure suppose node 1, wishing to send a message to a neighboring node, senses the channel idle and transmits its packet. It is clear in this network, 1's packet will be successfully received only if node 4, as well as its destination, are not transmitting. Thus we can conclude, that in an optimized network (where the objective is to maximize the number of successful transmissions in the network over time) there is some relationship between the rates at which nodes in the network present packets to an idle channel. In particular, for our example, the rate that node 1 presents packets to an idle channel in an optimized network, is functionally dependent upon the rate that node 4 presents packets to the channel.

Another relationship can be hypothesized by supposing that node 1, wishing to send a packet, senses the channel busy. In single-hop networks this implies that node 1 will refrain from transmitting until a future time. In our example, however, if node 1 senses the channel busy due to a transmission from node 3, this implies (assuming that node 1 is not sensing the channel if it is receiving a message) that node 3 is transmitting a message to node 4. Thus if node 1 transmits to node 2, its reception will be successfully received without interrupting node

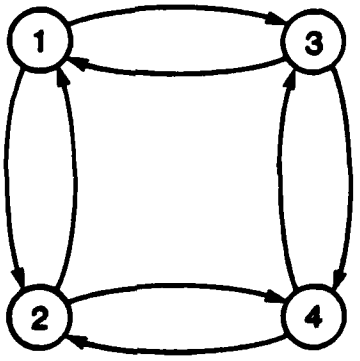


FIGURE 5.1
A sample network.

3's transmission. Thus it appears that if node 1 senses the channel busy it should not necessarily refrain from presenting packets to the channel. Clearly the rate at which it offers such packets will depend upon the topology and traffic generation rates of the nodes in the network but we can conclude from these motivating arguments that, in some cases, throughput would be increased if nodes transmitted even when the the channel was sensed busy. This characteristic, namely transmitting even though other nodes are using the channel, explains the appellation "rude-CSMA" for the protocols we define in the following section. We should mention that these relationships were also discussed in Chapter 1 when we showed in Figure 1.1 that there were cases when it was advantageous for transmitters t_1 and t_2 to transmit even if they sensed a busy channel.

This hypothetical reasoning leads us to question how the rates of offered packets to the channel should be adjusted to achieve maximal throughput. In this chapter we investigate a family of protocols which attempt to use the local information contained in the state of the channel to determine the rate at which nodes transmit packets on the channel. The question we hope to shed light upon in this chapter is: How can the "free" information obtained from sensing the channel be best used to maximize the throughput of multi-hop networks. We will see that the family of protocols we create to answer this question include both ALOHA and CSMA as special cases, and, being generalizations of them, should have an optimized performance that is at least as good as either of them. To further motivate the protocols that will be precisely defined in the next section, suppose node i is ready to transmit a packet to one of its neighbors. After sensing the channel, suppose i finds the channel busy. Besides the binary information contained in the channel status, i also obtains the power of the received signal. Suppose then (Figure 5.2) that the received power is very small. This implies that the number

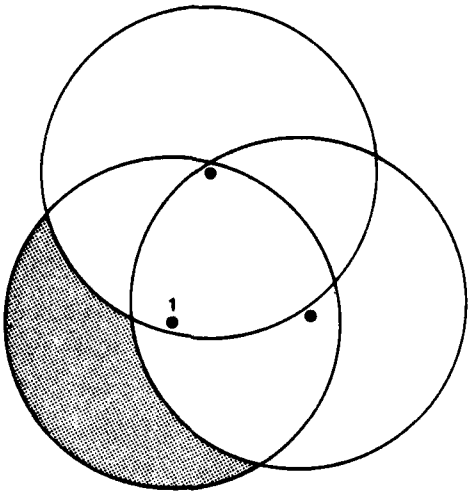


FIGURE 5.2
Small Received Power.

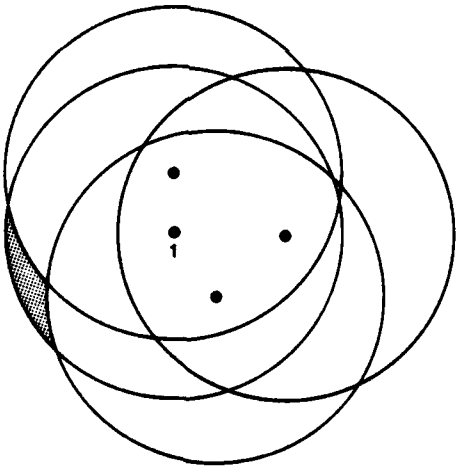


FIGURE 5.3
Large Received Power.

of transmitters within range of i is small and furthermore that they are not very close to i. If we assume that all nodes transmit with the same power, this implies that the immediate area surrounding i is free of transmitters and thus that there a possibility that if i transmits, the signal will be successfully received by its destination. The percentage of area of the shaded region of the circle shown in the figure is in fact the probability that i will be successful. On the other hand, suppose the received power is large (Figure 5.3). This implies that either there are a few transmitters very close to i or many far away. In the first case the environment covered by a transmitter very close to i would greatly overlap that covered by i and thus a transmission by i would most likely cause a collision. In the second case, Figure 5.3, since there are a large number of transmitters around i, the probability that one of i's neighbors hears a silent channel is very small and thus if i transmits, its signal would have a high probability of causing a collision. We can conclude from this hypothetical scenario that the decision to transmit depends upon the number of transmitters and their locations in i's immediate environment. In the protocols defined in this chapter we will assume that given the power of the received signal on the channel, the number of transmitting nodes within the hearing distance of any given node can be estimated. Later we will show that this assumption, although being unrealistic, is not too restrictive.

5.2 The Protocol

We will define the state S of the network to be a binary vector $S = (s_1(S), s_2(S), \cdots, s_n(S))$ where n is the number of nodes in the network and $s_i(S) = 1$ if node i is transmitting and 0 otherwise. For a given node, say node i, there exists a subset of the other nodes of the network which are its neighbors. This set is denoted by

$$A_i = \{ j \mid \text{node } j \text{ is within hearing distance of node } i \}$$

In general for a multi-hop network, A_i is a proper subset of all the nodes in the network. Thus i's decision to transmit will depend only upon the states of the nodes contained in its hearing region A_i. For a given state S, suppose node i has a packet which is ready to be transmitted. Node i senses the channel and based upon the power of the received signal, estimates the

number of transmitters within its local hearing distance. Denote this number by:

$$N_1^i(S) = \sum_{j \in A_i} s_j(S)$$

Likewise the number of neighbors not transmitting in i's hearing range can be estimated as:

$$N_0^i(S) = |A_i| - N_1^i(S)$$

We assume in our mathematical model that these estimates are exact. According to these values node i adjusts the rate at which it presents packets to the channel, and in particular, for state S, the rate at which node i transmits (assuming that $s_i(S) = 0$) is given by:

$$r_0^i(S) = \gamma_0 \, x^{N_0^i(S)} \, y^{N_1^i(S)} \tag{5.1}$$

where x and y are given parameters of the protocol and γ_0 is a given arrival rate of packets to each node of the network (from the attached host computer). In the case that node i is a transmitter, $(s_i(S) = 1)$, we assume the packet transmission time is exponentially distributed with an average length of $1/\mu$ time units and thus the rate at which i's transmission stops is given by:

$$r_1^i(S) = \mu \tag{5.2}$$

For any given state S, with these rate definitions, one can determine the rate at which the system changes into any other state S', and it is clear that the state vector and rate definitions define a continuous time, finite state, Markov process. A generalized form of this process can be found in [Kell79a]. Observe that this two parameter family of protocols contain ALOHA (x=1, y=1) and CSMA (x=1, y=0) as special cases. To optimize the performance of the network we need to determine the (x, y) values that maximize the expected number of successful transmissions over a unit time interval. For a given state S, define $E(S)$ to be the expected number of successful transmissions for that state. It is clear that $E(S)$ can be calculated if the traffic matrix, topology, and capture ratio for the system are known. For example, suppose in Figure 5.4 that nodes are equally likely to transmit to any of their neighbors. Nodes in Figure 5.4 are labeled 1 if they are transmitters and 0 otherwise. We can determine $E(S)$ by calculating the probability that silent nodes in the network successfully receive a packet destined towards them. Thus node 2 in the figure is adjacent to only one transmitting node (node 1) which transmits in that direction with probability 1/4. The expected success then for node 2 is 1/4 as it is for nodes 3 and 4. In the same manner nodes 6 and 7 have a 1/4 probability of receiving node 5's transmission successfully, whereas node 8 has no chance of receiving either

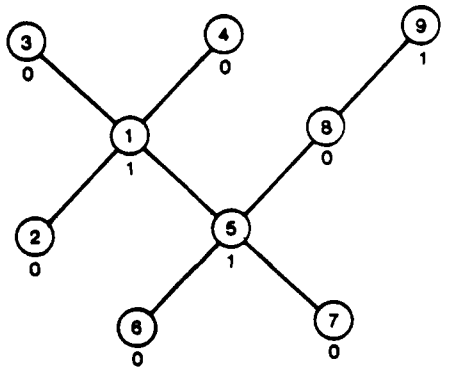


FIGURE 5.4
Calculating E.

node 5's or node 9's reception since they will collide. For the state shown then, the expected success is:

$$E(S) = 3\left(\frac{1}{4}\right) + 2\left(\frac{1}{4}\right) = \frac{5}{4}$$

Similar procedures can be generated that handle cases with different assumptions of the traffic matrix and capture parameter than those of the previous example. Observe that these assumptions correspond to a packet radio network in which terminals have only a transmitter buffer, and radio units are assumed to capture the first transmitted signal they receive. This *perfect time capture* is an ideal assumption and could alternatively be modeled as an upper bound on throughput to a packet radio network without time capture in which the exponential message length corresponds to a bulk of very small packets. To explain this suppose a radio unit i is successfully receiving a transmission from another unit j over the time period $[0,t]$. It is possible that another node k, within i's hearing radius but not within j's range, starts transmitting at time t. Clearly, assuming non-capture, this implies that i will hear a collision at t, and will not successfully receive the rest of j's message if j was sending a single packet message. However if we assume that j was sending a series of very short packets, then only those sent after time t will be lost. In the limit as the packet size goes to zero, $E(S)$ then will measure the fraction of time the channel that is being used successfully when the system is in state S.

Let $\Pi(S,x,y)$ be the steady state probability of state S for a given (x,y). In Appendix C we show that with rate equations (5.1) and (5.2), $\Pi(S,x,y)$ is given by:

$$\Pi(S,x,y) = C\,\rho^{M(S)}\,x^{-B_0(S)}\,y^{B_1(S)}$$

where
$$M(S) = \sum_{i=1}^{n} s_i(S)$$
$$C = \Pi(0,x,y)\,x^{-B_0(0)}$$
$$\rho = \gamma_0/\mu$$
$$\Pi(0,x,y) = \left[x^{-B_0(0)} \sum_{S} \rho^{M(S)}\,x^{-B_0(S)}\,y^{B_1(S)}\right]^{-1}$$

In this equation C is a normalization constant, and for state S, $M(S)$ is the number of transmitters, $B_0(S)$ is the number of adjacent nodes that are not transmitting, and $B_1(S)$ is the number of adjacent transmitters. Using this we can write the expected number of successes in the network over a unit time interval as:

$$E(x,y) = \sum_{S} \Pi(S,x,y)\,E(S)$$

Naturally we would like to maximize this function over feasible (x,y) values. Besides the non-negativity of x and y however, there is a constraint concerning the average rate at which a node presents packets to the channel. This actual rate must not be greater than the arrival rate of packets to that node, which is defined to be γ_0. Rate equation (5.1) shows that this actual transmission rate is a function of S, and this, averaged over all states, must therefore be less than γ_0. We thus have, letting $\bar{s}_i(S)$ be the complement of the i^{th} component of S:

$$\sum_{S} \gamma_0\,x^{N_0^i(S)}\,y^{N_1^i(S)}\,\bar{s}_i(S)\,\Pi(S,x,y) \leqslant \gamma_0 \qquad i = 1,2,\cdots,n$$

For a given topology and traffic matrix, we can thus formalize the mathematical program **P** as:

Program P

$$\max_{(x,y)} E \triangleq \sum_{S} \Pi(S,x,y)\,E(S)$$

subject to:

$$1 - \sum_{S} x^{N_0^i(S)}\,y^{N_1^i(S)}\,\bar{s}_i(S)\,\Pi(S,x,y) \geqslant 0 \quad i = 1,2,\cdots,n$$

$$x \geqslant 0$$

$$y \geqslant 0$$

This is the problem that we will address in this chapter. We should note however that optimal

(x,y) values are functions of the topology and traffic matrix of the network. We will call the first constraint equation in **P** the *flow constraint* and will say that it is saturated if the equation is an equality for some equation.

5.3 Discussion of Results

In this section we will discuss the results we obtained when **P** is optimized over different topologies. We wrote computer programs which, for a given topology, calculated $E(S)$ and $\Pi(S,x,y)$ and then optimized **P** over all feasible (x,y) pairs. Besides special topologies which we studied, we also ran **P** for connected networks which were randomly generated for varying mean densities. Unless stated otherwise, in all networks we have assumed a uniform local traffic matrix. As typical of most Markovian models, the state space of the system grows exponentially with the number of nodes of the network, and thus to keep the program computationally tractable, we restricted our optimization to relatively small networks. We do not believe however, the applicability of our results are dependent upon the size of the network.

5.3.1 Special Topologies

In figure 5.6 we plot the values of x and y that achieve optimal performance for the network of figure 5.5, which is a generalized version of the network used to motivate this work. The curves for the graph in figure 5.1 show similar behavior. These curves have many interesting properties. First we observe that over the range of ρ shown there are three distinct types of behavior. For very small ρ values, $\rho \leqslant .15$, we see that x and E increase rapidly while y remains very small. The increase of E over this range is explained by the fact that for small ρ values, there are very few transmissions, and thus very few which cause collisions (or interpreted in the case of perfect time capture, few transmissions that aren't first) and thus increasing ρ tends to increase E in a linear manner. We see this behavior clearly as ρ goes from 0.05 to 0.1 during which E doubles from about 0.35 to about 0.6. As ρ continues to grow beyond 0.1 the increase is less than linear due to some collisions in the network. It is clear that over this range, since there are so few transmissions in the network, the y parameter of rate equation (5.1) is non-critical. Using an interactive program we wrote to determine how E varied as a function of y, we found very little change over all values of y, thus the sudden increase in y at $\rho \cong .2$ should not be interpreted as demonstrating singular behavior.

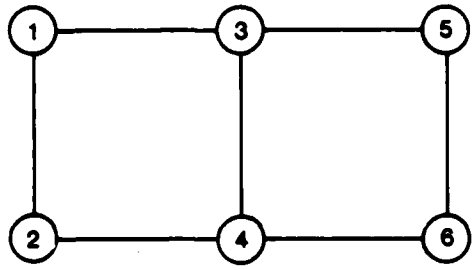


FIGURE 5.5
A Lattice Network.

The increase in x over the range $0.05 \leqslant \rho \leqslant 0.2$ can be explained by first observing that for this range the chance of collisions is small and thus nodes present packets to the channel as fast as they can generate them, which in our case is approximately equal to $\gamma_0 x^2$. We should mention that for these small values of ρ, the optimal x value lies on the boundary of the feasible region, hence the flow constraint of **P** is saturated. We can explain the linear behavior of the x curve over this region with an intuitive argument. From our previous discussion we know that E is linear in ρ for small values, and it can be approximated closely by $E = 4.6\rho + .1$. For the most part, successful states for these small ρ values correspond to one transmitter in the network. Thus using our equation for $\Pi(S,x,y)$, we can write $E = k\rho x^{-2.33}$, where k is a constant of proportionality. We can then write:

$$4.6\rho + .1 = k\rho x^{-2.33} \tag{5.3}$$

Using the initial values from figure 5.6 we can determine $k = 10.09$. Solving equation (5.3) for x therefore yields:

$$x = \left(\frac{10.09\rho}{4.6\rho + .1}\right)^{1/2.33}$$

which for small values of ρ is very closely approximated by:

$$x \cong 2\rho + 1.1$$

This then explains the linear shape and slope of x over this region. In summary then, over the first region of the curve in figure 5.6, E and x are approximately linear with slopes 4.6 and 2 respectively, and y is not a critical parameter of the system.

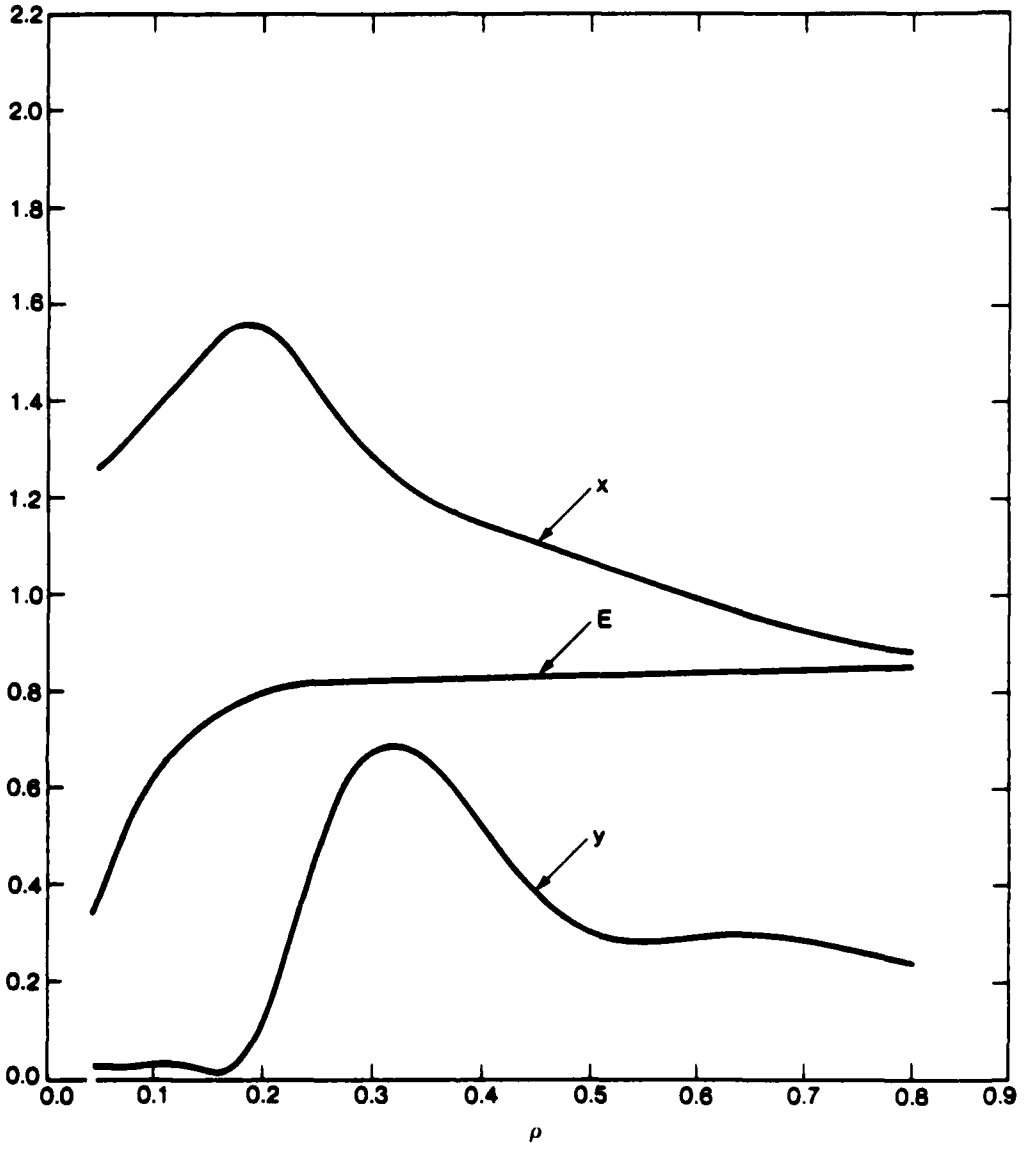


FIGURE 5.6
Curves for the Lattice Network.

The second region of the graph, from ρ=0.2 to ρ=0.35, will now be explained. First we observe that the expected success rises only slightly over this range. Although there are more transmissions in this range, there are also more collisions which limit the number that are successful. The sudden increase of y at $\rho \cong 0.2$, as mentioned previously, should not be looked on as being as singular, but does demonstrate the increased importance of y's effect on the throughput of the system. Observe that over this range $y<1$, and thus from equation (5.1), this parameter acts to inhibit transmissions when there is a neighboring transmitter (although not to the extent of excluding such transmissions). This supports our conclusion from the scenario in the beginning that sometimes in a multi-hop network it is beneficial to transmit even if the channel is sensed busy. The decrease of x over this range arises because it is no longer on the boundary of the feasible region. However, since x is greater than 1 in this region, we conclude that this parameter tends to increase the rate of packets offered to the channel during idle channel periods.

Over the region $\rho>0.35$, E is again not altered to any great measure. Both x and y decrease now in an effort to prevent collisions on the channel. As ρ becomes much larger, we see that $x<1$, which indicates that even the x parameter tends to act to inhibit the number of packets on the channel. Over this region, the flow constraint equation is not saturated for any node in the network indicating that the effective transmission rate of packets on the channel is less than γ_0. Hence optimizing the throughput of the network has the effect of limiting the flow of packets in the network. We will explain the general shape of the x curve when discussing random topologies in the next section.

The next special topology we will study is that of ring networks as shown in figures 5.7 and 5.10. In each of these rings we have assumed that messages always travel clockwise (thus the arrows in the figures). We observe that y=0 for all ρ values and thus nodes in such networks should refrain from transmitting if they sense a busy channel. This is actually obvious from the defined topologies since the sensed transmission would either be directed towards the sensing node or be the intended recipient of the sensing nodes transmission. In each case then transmitting would cause a certain collision. Knowing that the nodes will use the CSMA protocol allows us to solve for the steady state throughput using an alternative method, thus giving intuitive insight into the solution and corroborating our equation for $\Pi(S,x,y)$. To generate the states for this process, we observe there are five states which have one transmitter, and five states that have two transmitters (we do not allow two transmitters to be adjacent). We can thus form the Markov chain shown in figure 5.9. In this figure α is the arrival rate of messages

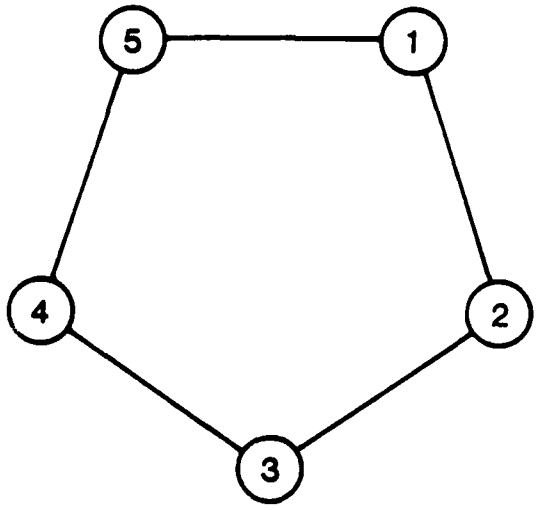


FIGURE 5.7
A Ring of Five nodes.

into each node and corresponds to $\gamma_0 x^2$ in our model. Solving this chain yields:

$$\begin{aligned} P_0 &= 1/(1+5(\alpha/\mu)+5(\alpha/\mu)^2) \\ P_1 &= 5(\alpha/\mu)P_0 \\ P_2 &= 5(\alpha/\mu)^2 P_0 \end{aligned}$$

where P_i is the equilibrium state probability of state i. It can easily be checked that using our equation for $\Pi(S,x,y)$ yields the same set of equations with $\alpha = \gamma_0 x^2$. In each of states 1 and 2 there is an expected success of 1 packet, and thus maximizing the throughput of the network corresponds to maximizing the probability that the system is in states 1 or 2. This can most easily be performed by minimizing the probability the system is in state 0. It is clear that this occurs for $(\alpha/\mu) \to \infty$, and in figure 5.8 we see the monotonic rise of throughput with ρ, and the increase in x which corresponds to lying on the boundary of the feasibility region.

Quite a different phenomena is observed for the case of a ring with an even number of nodes as shown in figure 5.10. To intuitively motivate this we draw the corresponding Markov

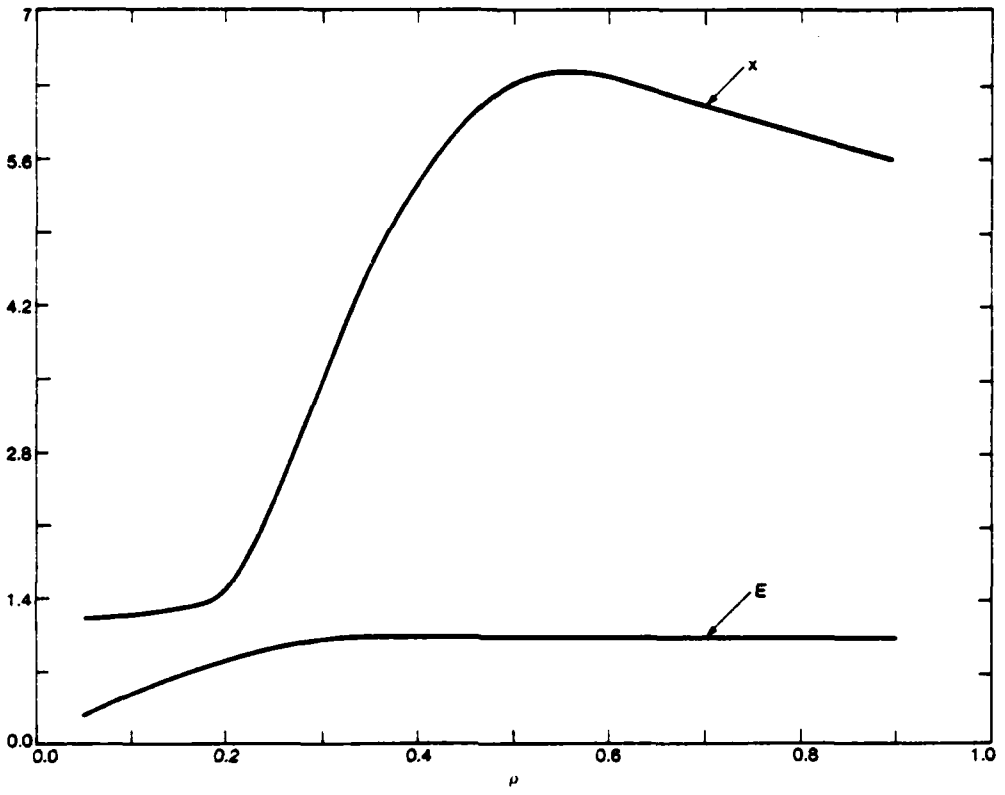


FIGURE 5.8
Curves for the Five Node Ring (y=0).

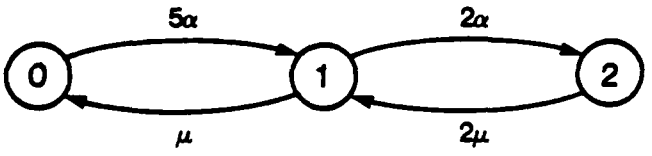


FIGURE 5.9
Markov Chain Representation of a 5 Node Ring.

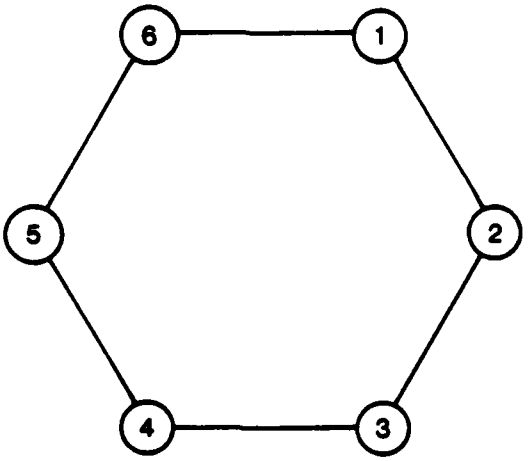


FIGURE 5.10
A Six Node Ring.

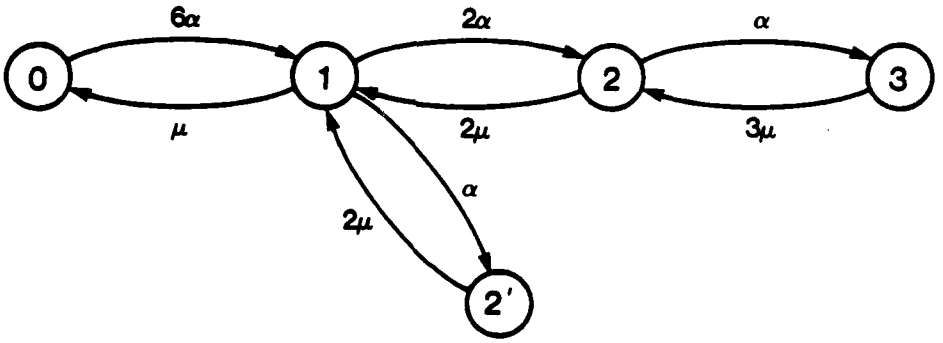


FIGURE 5.11
Markov Chain Representation of a Six Node Ring.

chain in figure 5.11, which can be solved to yield:

$$
\begin{aligned}
P_0 &= 1/(1+2(\alpha/\mu)^3+9(\alpha/\mu)^2+6(\alpha/\mu)) \\
P_1 &= 6(\alpha/\mu)P_0 \\
P_2 &= 6(\alpha/\mu)^2P_0 \\
P_{2'} &= 3(\alpha/\mu)^2P_0 \\
P_3 &= 2(\alpha/\mu)^3P_0
\end{aligned}
$$

where $\alpha=\gamma_0 x^2$ and state 2 corresponds to a configuration congruent to 1 and 3 being transmitters and state 2' to 1 and 4 being transmitters.

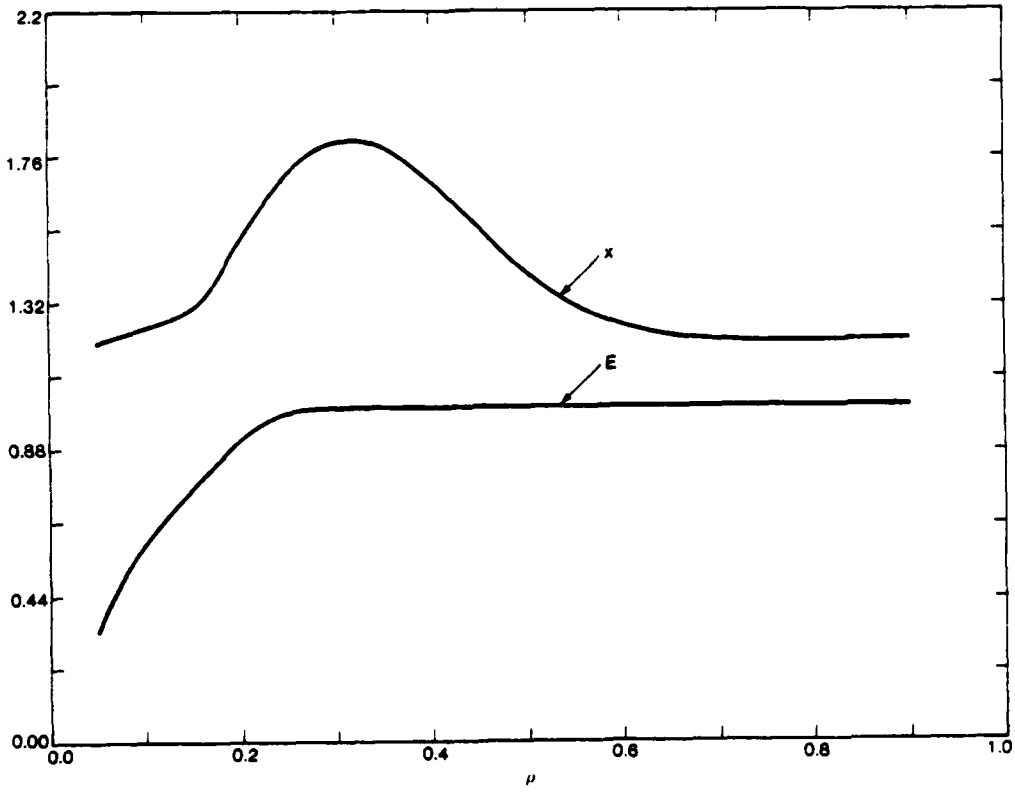


FIGURE 5.12
Curves for the Six Node Ring ($\gamma=0$).

The expected success for each state is: 0 for states 0 and 3, and 1 for state 2, and 2 for state 2'. Thus we wish to maximize $P_2+2P_{2'}$ over α. This situation is much different than the previous odd number ring because increasing α has the effect of increasing P_3 which has an expected throughput of 0. The objective then is:

$$\max_{\alpha} \quad \frac{6(\alpha/\mu) + 12(\alpha/\mu)^2}{1 + 2(\alpha/\mu)^3 + 9(\alpha/\mu)^2 + 6(\alpha/\mu)}$$

This maximum occurs for $(\alpha/\mu) = 1$ at which point the objective function is equal to 1. In figure 5.12 we see this behavior in the curves for x and E. For $\rho \leqslant .3$ the flow constraint in **P** is satisfied and the (α/μ) values are much less than the optimal value of 1 because there is not enough input to satisfy this condition. For $\rho > 0.3$ however, the input into the system is sufficient to reach optimal throughput and the x values shown act to control the rate of this input to achieve optimal throughput. For example for $\rho = 0.5$, $x = 1.4$ yields $\alpha = \rho x^2 = 0.98$ (where we have normalized μ to be equal to 1). We notice over this range that the optimal throughput of 1 packet per time unit is obtained.

We observe with rings a striking difference in the behavior of the system (in terms of the $\alpha = \gamma_0 x^2$ parameter) which depends on the number of nodes in the network. In particular for an odd number of nodes, optimal performance was obtained for $\alpha \to \infty$, whereas for an even number of nodes we saw that $\alpha < \infty$ achieved maximal throughput. This dependency results from combinatorial considerations on how transmitters could be distributed in the nodes such that no two adjacent nodes were transmitters. In odd node rings, when the maximum number of transmitters is generated, the expected throughput for the configuration is greater than zero (in fact it is always equal to 1), whereas for even node rings, this configuration yields a zero throughput. This tends to make optimal α values for odd node rings greater, but not necessarily infinite, than those for even node rings. We might question if such behavior is obtained when the rings are cut at one spot and made into tandems.

FIGURE 5.13
A Five Node Tandem Network.

In figures 5.13 and 5.15 we show a five and six node tandem with their associated curves in figures 5.14 and 5.16. Both graphs show similar behavior. Once again for $\rho < 0.4$ the flow constraint equation in **P** is saturated and after that the x parameter acts to meter the actual flow of packets into the network.

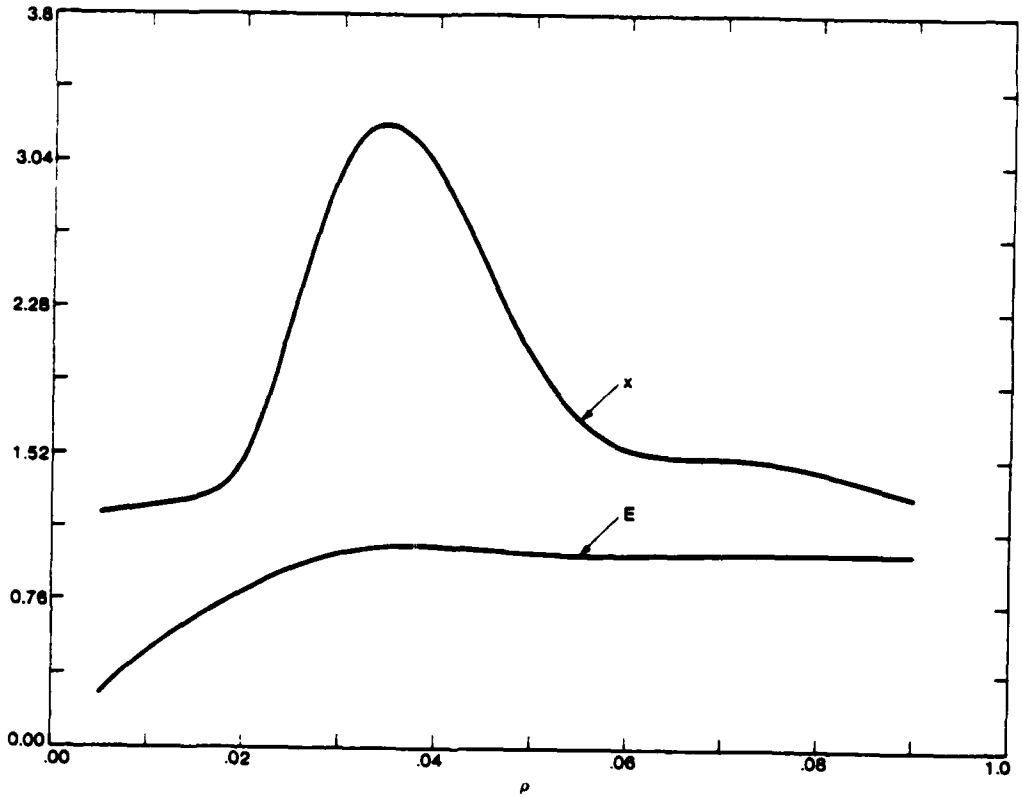


FIGURE 5.14
Curves for the Five Node Tandem (γ=0).

FIGURE 5.15
A Six Node Tandem.

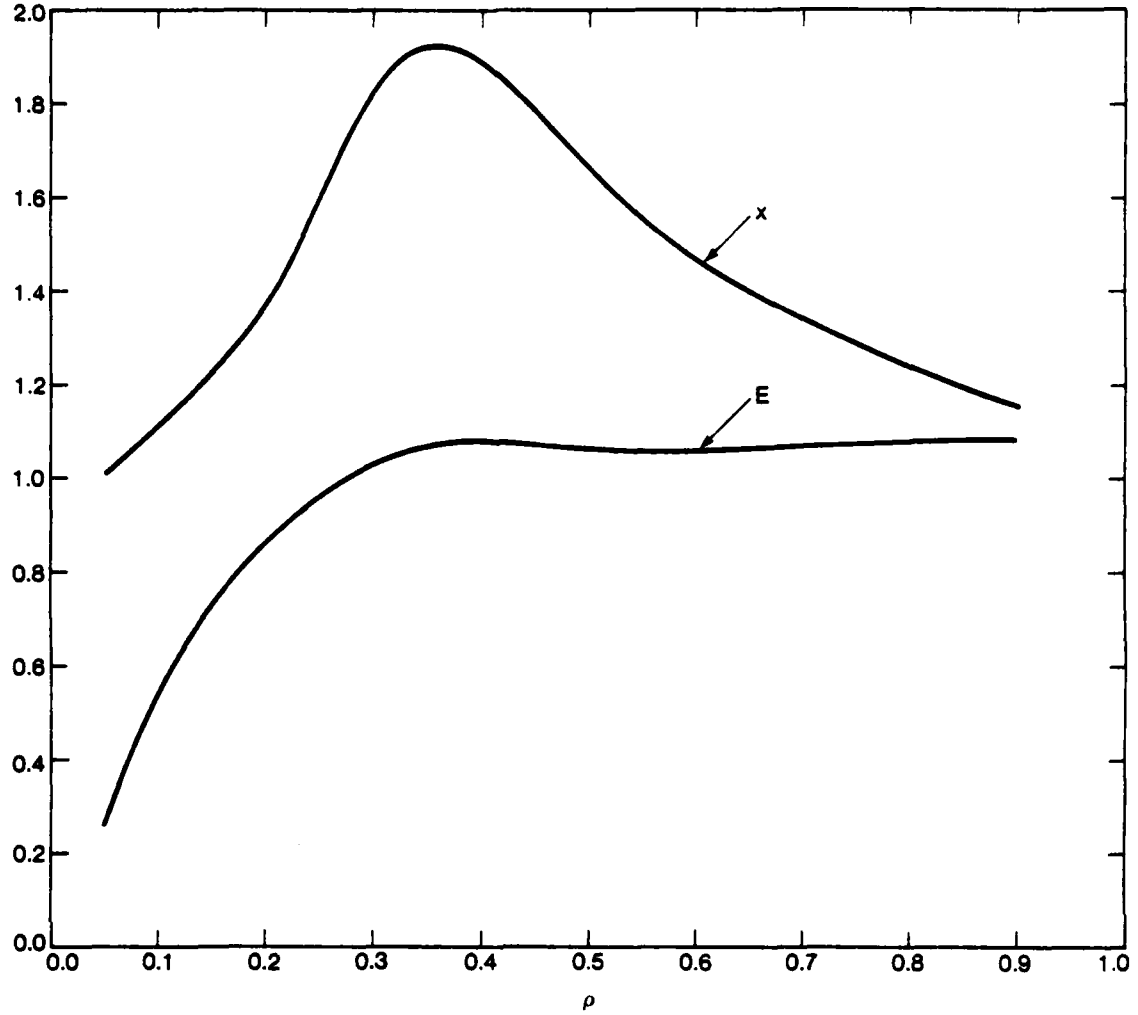


FIGURE 5.16
Curves for the Six Node Tandem (y=0).

5.3.2 Random Topologies

In this section we report on the results which were obtained when **P** was run on graphs which were randomly generated. Many such graphs were created, with varying mean densities and all such graphs exhibited similar behavior. This characteristic allows us to make the blanket statement that in a random multi-hop network, over the continuum of protocols defined by all possible (x, y) pairs, optimum performance was obtained for those pairs that had y=0. As stated in the introduction this corresponds to a CSMA type protocol. From the many graphs which were analyzed we have picked two to show here. This selection was based more on the aesthetic qualities of the graphs rather than any particular characteristic of the curves associated with solving **P**. In figure 5.17 we show a 7 node network with a mean number of neighbors, N equal to 2.28. The corresponding set of curves are shown in figure 5.18. We see that for low values of ρ, both E and x are linear indicating that there are very few collisions in the network. The slope of E becomes nearly zero for $\rho > 0.2$ implying that after this point the amount of new successful transmissions in the network is small in comparison to the number of

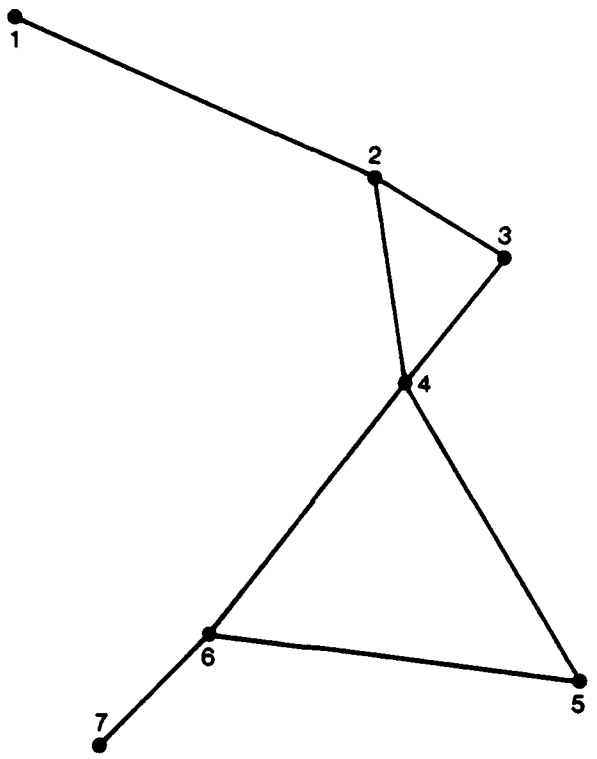


FIGURE 5.17
A random graph with $N = 2.285$.

collisions that arise from the increased traffic load. The shape of the x curve for $\rho > 0.3$ can be explained by first noting that over this range the flow constraint of **P** is not saturated. Next we make a definition. Suppose for each node i, we investigate all state vectors having the property that i and all its neighbors A_i are silent. Under these conditions, if i senses the channel, it will detect that it is idle and offer packets to the channel at a rate defined by the protocol. Let I_i be the set of all state vectors satisfying this condition. We can then define the *effective number of idle neighbors*, denoted by N^*, by finding the average number of silent neighbors a randomly selected idle node has when it senses the channel prior to a transmission. This can be written as:

$$N^* = \sum_i |A_i| \sum_{S \in I_i} \Pi(S, x, y)$$

Because of topological variations, some nodes are more likely to be both idle and surrounded by idle neighbors than others (in fact the probability is decreasing with the number of neighbors) and thus in general $N^* \neq N$. Define f to be the average time a random node in the network waits before transmitting when it hears an idle channel. Since E does not vary much over

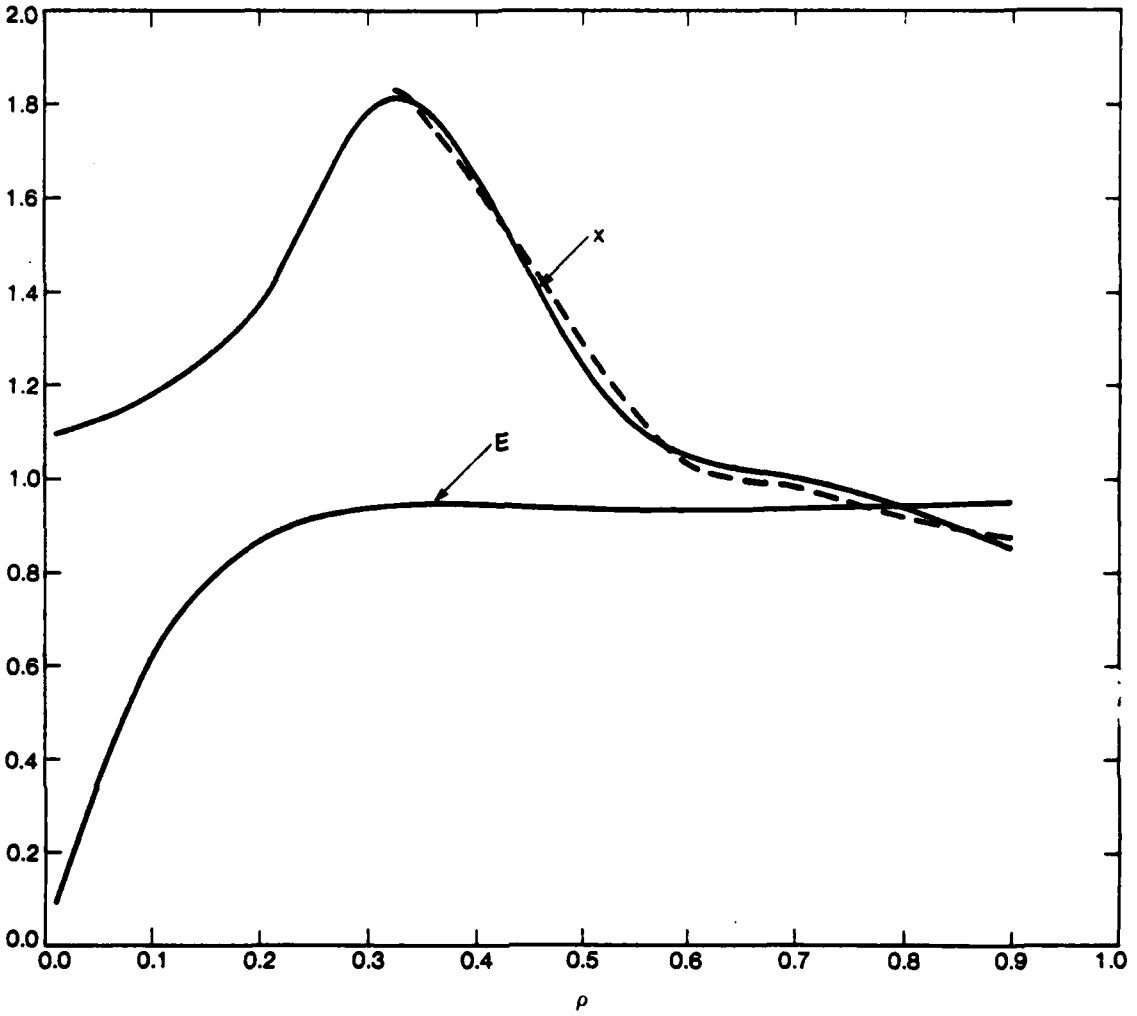


FIGURE 5.18
Curves for graph in figure 5.17 (y=0).

ρ>0.3 it is not surprising to find that x varies as a function of ρ to preserve the fraction of time nodes transmit when sensing an idle channel. This has been verified by observing (by using an interactive program) that the probability distribution for state vectors does alter significantly as ρ ranges greater than 0.3 for optimized x values. We thus can equate rates to obtain:

$$\rho x^{N^*} = 1/f$$

We can solve for f and N^* by picking two points from figure 5.18. At ρ=0.72 we have x=1

and thus can write $0.72=1/f$ or $f=1.36$. At $\rho=0.32$ we have $.32(1.8)^{\mathcal{N}^*}=.72$ which implies that $\mathcal{N}^*=1.4$. Thus we can write x as a function of ρ as:

$$x = (1/f\rho)^{1/\mathcal{N}^*} = (.72/\rho)^{-1} \quad \rho>0.3 \tag{5.4}$$

Equation (5.4) was used to generate the second curve in figure 5.18 and we see a relatively close match. In general then, for a CSMA environment, we can write a general equation relating x with ρ as:

$$x = (1/f\rho)^{1/\mathcal{N}^*}$$

Intuitively, for the graph of figure 5.17, we would expect $\mathcal{N}^*<\mathcal{N}$ since nodes having the greater number of neighbors, node 4 for instance, have a much smaller probability of being both idle and surrounded by idle neighbors, than a node, say node 1, that has far fewer neighbors.

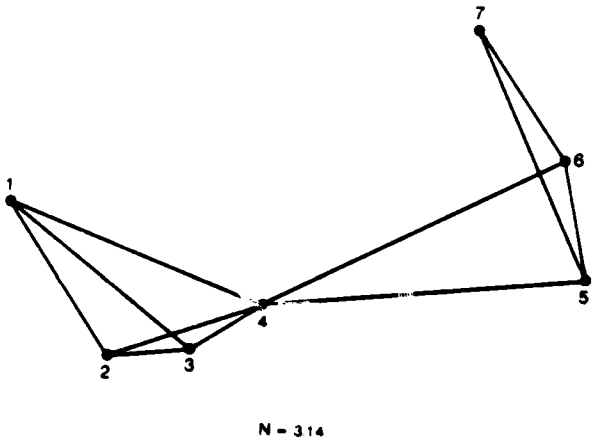


FIGURE 5.19
A random graph with $\mathcal{N}=3.14$.

In the second graph of this section, figure 5.19, we again see in its associated set of curves, figure 5.20, similar behavior. As stated in the beginning of this section, this behavior was typical of all the curves for the random graphs we generated.

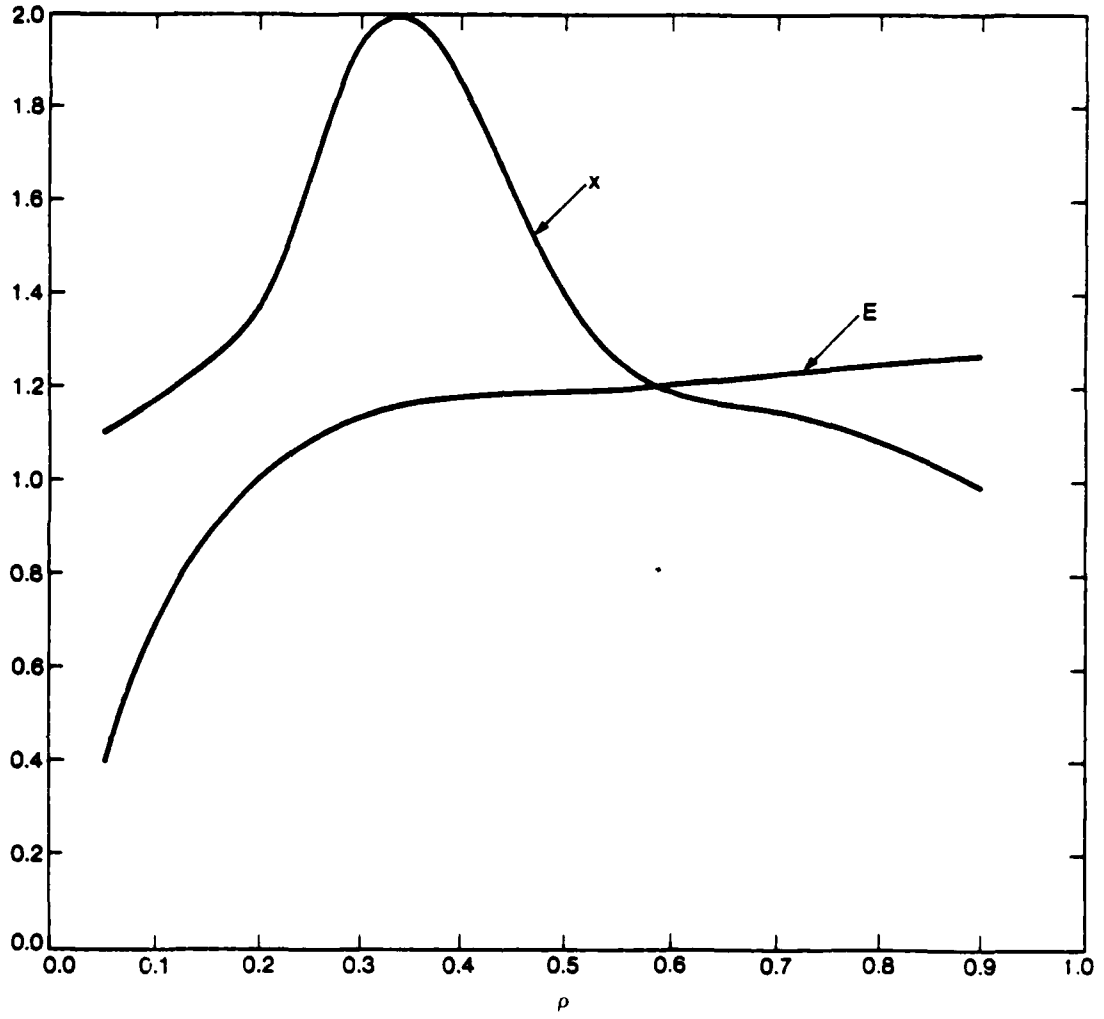


FIGURE 5.20
Curves for figure 5.19 (γ=0).

5.4 Conclusions

In this chapter we have defined rude-CSMA which is a two parameter continuum of protocols for multi-hop channel access. This family of protocols has been shown to include both ALOHA and CSMA as special cases. The optimum performance of rude-CSMA for special topologies was investigated, and we have seen that for lattice type networks, under certain conditions, transmitting even after sensing a busy channel tends to increase the throughput of the system. In all other topologies studied including rings, tandems, and random graphs, optimal performance was obtained when the protocol used was CSMA with optimized channel input rates. Thus we must come to the interesting conclusion that for practical networks, rude-CSMA is not that "rude" afterall. This justifies the statement in the introduction that assuming nodes could determine the number of transmitters from the received power of sensed signals was not too restrictive an assumption. The binary information contained obtained from sensing the channel appears to be sufficient to determine the optimal transmission policy. The curves generated for these protocols under varying assumptions have proved to be intuitively insightful and their interesting properties have been explained.

CHAPTER 6
CONCLUSIONS

In the previous chapters we analyzed the interference problem under various sets of policy state information. We have also seen that the difficulty of the problem concerned the fact that all nodes of a packet radio network are interrelated, and that changing the behavior of one node in such a network, causes global effects. Coupled, dependent systems like these are very difficult to analyze and our approach in each of the models we analyzed, undertook in some manner a means to control the dependency. In Chapter 2 we essentially ignored the dependency and determined the best performance obtainable for slotted-ALOHA networks in which only the mean density of terminals on the plane was known. This model corresponded to solving the interference problem using a minimal set of policy state information, and in Chapter 3 we established an upper bound on the performance for all protocols which use such information. In Chapter 4 we strictly controlled the dependency between nodes of the network by using the global information of their location and traffic characteristics. We saw the strong inter-relationship between all of the nodes when we attempted to determine optimal time frames for networks using spatial-TDMA, and needed to resort to solving the capacity assignment problem over the class of frames that were randomly generated. The coupling between the nodes of the network was again encountered in the fifth chapter where we sought to control the inter-dependencies by determining a protocol, rude-CSMA, which used only local information to determine when to transmit packets on the channel. The coupling in this model was restricted, for each node, to the subset of nodes in its local environment. The particular rate definitions given for the protocol allowed such a de-coupling and also allowed us to solve for the steady state probabilities of the system state in closed form.

We may pose the following question: Where does the subject go from here? It is the author's opinion that probably the most important area of research, that would shed light on the interference problem as well as numerous other such such problems, concerns the development of more sophisticated mathematical tools which are especially designed to work with dependent systems of interacting processes. Most of the relevant mathematics used to analyze such systems, such as stochastic processes and probability theory, are only tractable for systems in which the random variables are independent. Short of having any other mathematical tools, the network analyst often makes the rather blatant assumption that the random variables associated with the network are independent. This assumption is certainly false. Although the approximations that arise from such assumptions often lead to models that are good predictors of network performance, without proper validation procedures there seems to be no mathematical way to determine the effect that certain assumptions have on the accuracy of the resultant

network model. This puts network designers in the rather nebulous position of not knowing what to believe, and forces them to simulate, or in some other way, to validate the model and thus nullify the "precision and cost effectiveness" of analysis. A current trend in solving coupled systems of this type consists in establishing sets of equations for the state of the system in terms of unknown parameters. These parameters form the coupling between state variables. The solution of these equations then involves guessing initial values for these parameters and then iterating through the equations until they converge to a final set of values. These final parametric values form a consistent set of equations for the system *but* do not necessarily correspond to an optimal solution of the system. Other methods for bypassing the coupling of the random variables, such as assuming average behavior in all relationships, suffer from the same flaws, namely, optimality is not easily shown and procedures are needed to validate the model. There is thus a great need for more sophisticated mathematical tools.

In pertaining to extensions and generalizations of this dissertation, we can make the following comments. In Chapter 2 we assumed that terminals were uniformly distributed on the plane in an effort to model a mobile packet radio network. In many networks, however, there is a tendency for nodes to form clusters rather than to be uniformly spread. For example, taxis tend to be more densely populated in areas of a city in which the demand for transportation is greatest. To model such a network, one would have to change the model of the second chapter to include the case where terminals were distributed on the plane according to a non-uniform density. This would imply that the range and probability of transmission would depend upon the local topology of the network. In Chapter 4 we solved the capacity assignment problem for spatial-TDMA networks over the class of frames that were randomly selected and claimed that this formed an upper bound to the mean delay obtained by an optimal frame. An important contribution would be an algorithm that efficiently produced an optimal assignment of time slots to the cliques of the network. The author of this dissertation worked intensely on this problem but was only successful in developing heuristic solutions that required substantial computation for their execution. Since spatial-TDMA could be a very useful protocol for local area radio networks, optimal performance is an important consideration. We commented in the conclusions to Chapter 4 that a form of spatial-TDMA that worked in a mobile environment would be desirable. Although the author has developed a formalization of such a protocol, more work needs to be done to make the protocol both efficient and practical. We can make similar comments about possible generalizations to rude-CSMA of the fifth chapter. In that model, solutions were obtained only for the case where the average rate of incoming packets to nodes of the network was the same. Once again this uniform assumption is often not found in practice (in the above example, taxis in the central part of the city may be using their transmitters more often than those on the outskirts). Unfortunately assuming different rates changes the tractability of finding a closed form solution to the Markov process defined by the state transition rates, and thus increases the complexity of finding optimal parametric values. If an efficient

method by which terminals in a mobile network determine their local topology (the locations of their neighboring nodes) can be developed, then an extension and marriage of rude-CSMA and spatial-TDMA could be possible. This, besides being interesting, certainly would increase the performance of the network.

APPENDIX A

An Expansion of the Integral of Chapter 2

We claim

$$\int_0^x t^2 e^{-kt^2}\,dt = \frac{xe^{-kx^2}}{2k} \sum_{j=1}^{\infty} \frac{(4kx^2)^j\, j!}{(2j+1)!} \tag{A.1}$$

Define

$$F(n) = \int_0^x t^{2n} e^{-kt^2}\,dt$$

We then wish to find F(1). To do this we first set up a recurrence relationship between $F(n)$ and $F(n+1)$. Integrating by parts we find

$$F(n) = \frac{x^{2n+1} e^{-kx^2}}{2n+1} - \int_0^x \frac{t^{2n}}{2n} (1 - 2kt^2) e^{-kt^2}\,dt$$

$$= \frac{x^{2n+1} e^{-kx^2}}{2n+1} - \frac{F(n)}{2n} + \frac{k}{n} F(n+1)$$

$$= \frac{x^{2n+1} e^{-kx^2}}{2n+1} + \frac{2k}{2n+1} F(n+1) \tag{A.2}$$

Using this relationship, we claim

$$F(1) = \sum_{j=1}^{i-1} \frac{(2k)^{j-1} x^{2j+1} e^{-kx^2}}{a_j} + \frac{2k^{i-1}}{a_{i-1}} F(i) \quad \text{for all } i$$

where $a_j = 1 \cdot 3 \cdot 5 \cdot \cdots \cdot (2j+1) = (2j+1)!!$. We prove the assertion by induction. It is true for $i = 2$ from the recurrence relation (A.2). Therefore assume it is true for $i = L$, then we

can show it is true for $i = L+1$ as follows: By assumption

$$F(1) = \sum_{i=1}^{L-1} \frac{(2k)^{i-1} x^{2i+1} e^{-kx^2}}{a_i} + \frac{(2k)^{L-1}}{a_{L-1}} F(L) \tag{A.3}$$

Using (A.2), the last term can be written as:

$$\frac{(2k)^{L-1}}{a_{L-1}} F(L) = \frac{(2k)^{L-1}}{a_{L-1}} \left[\frac{x^{2L+1} e^{-kx^2}}{2L+1} + \frac{2k}{2L+1} F(L+1) \right]$$

$$= \frac{(2k)^{L-1} x^{2L+1} e^{-kx^2}}{a_{L-1}(2L+1)} + \frac{(2k)^{L}}{a_{L-1}(2L+1)} F(L+1)$$

but $a_{L-1}(2L+1) = a_L$ so

$$= \frac{(2k)^{L-1} x^{2L+1} e^{-kx^2}}{a_L} + \frac{(2k)^{L}}{a_L} F(L+1) \tag{A.4}$$

We can put the first term of (A.4) into the sum of (A.3) to obtain

$$F(1) = \sum_{i=1}^{L} \frac{(2k)^{i-1} x^{2i+1} e^{-kx^2}}{a_i} + \frac{(2k)^{L}}{a_L} F(L+1)$$

and thus we prove that the expression is true for the $L+1$'st term. If we continue in this manner we finally obtain

$$F(1) = \lim_{L \to \infty} \sum_{i=1}^{L} \frac{(2k)^{i-1} x^{2i+1} e^{-kx^2}}{a_i}$$

$$= \sum_{i=1}^{\infty} \frac{(2k)^{i-1} x^{2i+1} e^{-kx^2}}{a_i} \tag{A.5}$$

Simplifying a_i we have

$$a_i = 1 \cdot 3 \cdot \cdots \cdot (2j+1)$$

$$= \frac{(2j+1)!}{2 \cdot 4 \cdot \cdots \cdot (2j)}$$

$$= \frac{(2j+1)!}{2^j j!}$$

and using this in (A.5) we finally obtain (A.1).

APPENDIX B

Calculating the densities for one, two, and three hops

We will first review the assumptions and definitions discussed in Chapter 3. Let points be distributed according to a Poisson process on the plane. Create a non-directed graph by joining points with an edge if the euclidean distance between them is less than or equal to one unit. For a randomly selected point, p_0, another point p is i hops away, if there exists a path from p_0 to p containing $i-1$ other points and no other path exists with fewer points. Assuming there exists a path between any two points of the graph, we derive in this appendix, the probability density functions for the euclidean distance between points that are separated by one, two, and three hops. In particular looking at figure B.1, we will derive the density functions for the variables r_1, r_2, r_3, and also θ.

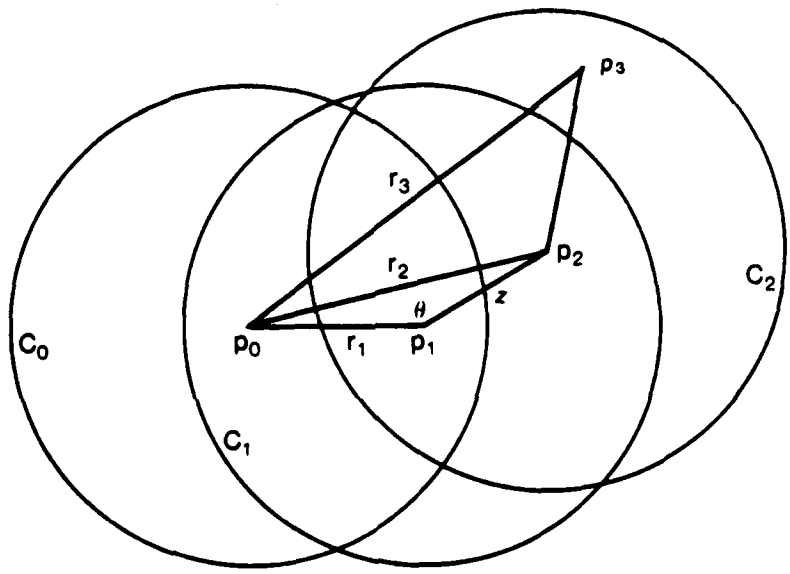


FIGURE B.1
A typical example of a three hop configuration.

We will use the following notations throughout the derivations (these are shown in figure B.1): p_i will be the i^{th} point from p_0, r_i will be the distance from p_i and p_0, z will be the distance

from p_1 and p_2, θ is the angle between the line segments (p_0, p_1) and (p_1, p_2), and C_i is the circle associated with p_i. We will assume that each circle has a radius of 1 unit. The random variables associated with the above quantities will be denoted as R_i, Θ, and Z. The derivations that are to follow require extensive use of the Law of Cosines and the Law of Sines for numerous defined triangles. To facilitate reading these derivations, we have followed the format of writing the initial equation that follows from these laws on the left side of the page and then showing the equation for the unknown quantity along the right side. It was felt in this way the triangle that was used in the derivation could more easily be identified by the reader. We will first derive the densities for R_1, R_2, and Θ producing conditional densities that will be necessary for the much harder derivation of the density for R_3.

B.1 The Density for R_1

Since points are randomly distributed according to a Poisson distribution, and since we are assuming that the resultant graph is connected, we see in figure B.2 that the probability distribution for the first hop can be formed by the ratio of the two areas:

$$P[R_1 \leqslant r_1] = \pi r_1^2/\pi = r_1^2$$

Differentiating this with respect to r_1 will give the probability density function:

$$h_1(r_1) = 2 r_1 \tag{B.1}$$

B.2 The Density for R_2

In figure B.3 we see that the conditional probability distribution can be written as:

$$P[R_2 \leqslant r_2 \mid R_1 = r_1] = B(r_1, r_2)/A(r_1, 1)$$

where $B(r_1, r_2)$ is the shaded area in figure B.3 and $A(r_1, r_2)$ is the area contained in C_1 that is not contained in C_0 or $B(r_1, r_2)$. It is clear that $B(r_1, r_2) = A(r_1, 1) - A(r_1, r_2)$, and thus we need to derive an expression for $A(\cdot)$. We will derive this function under more general conditions where the radii of the circles are not both equal to 1. We will use figure B.4 to facilitate this derivation.

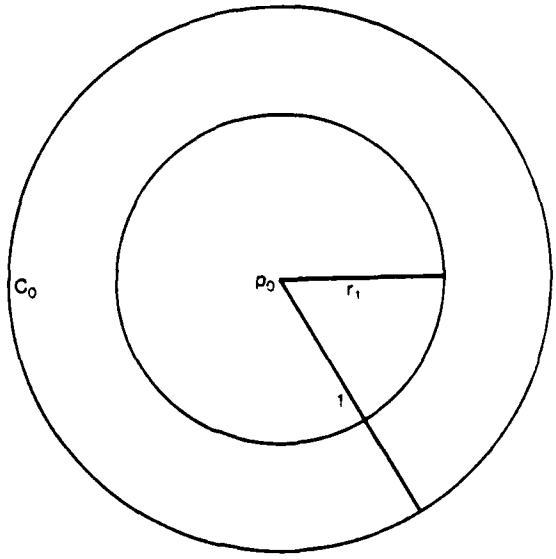


FIGURE B.2
Calculating the density for the first hop.

In this figure, $T(r_1, r_2)$ is the area formed by the triangle of vertices p_0, p_1, and P, $W(r_1, r_2)$ is the area which added to $T(r_1, r_2)$ would form the sector of radius r_2 and angle ϕ, and $A(r_1, r_2)/2$ is the shaded area in C_1. Since $A(r_1, r_2)/2$ is contained in a sector of radius 1 and angle ψ, we can write:

$$A(r_1, r_2)/2 = \psi/2 - W(r_1, r_2)$$

but $W(r_1, r_2)$ can easily be seen to be:

$$W(r_1, r_2) = \phi\, r_2^2/2 - T(r_1, r_2)$$

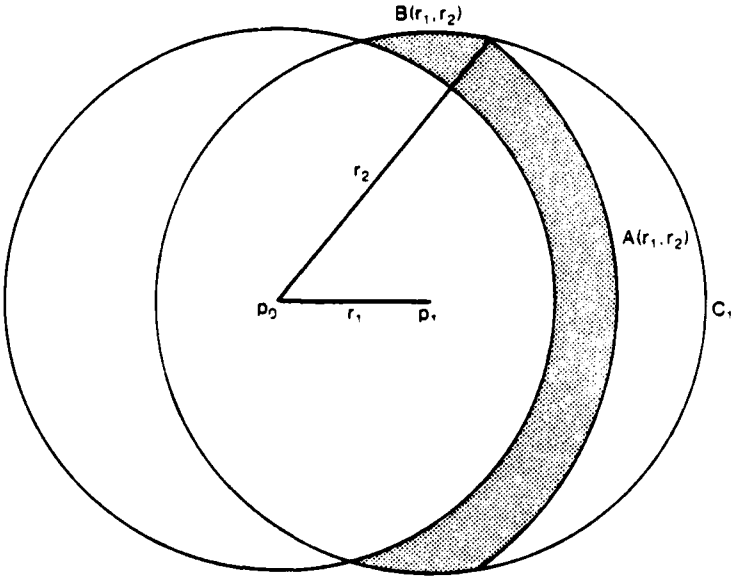


FIGURE B.3
Calculating the density for the second hop.

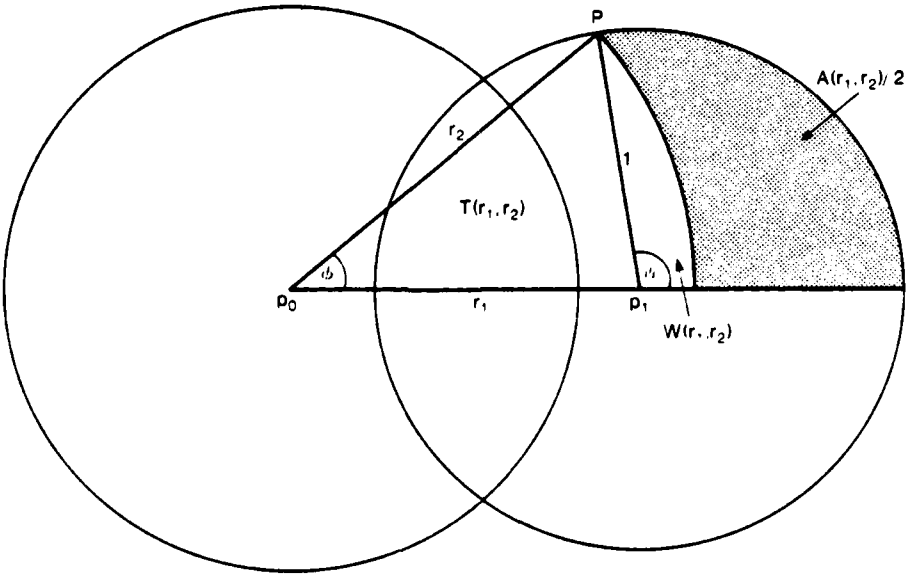


FIGURE B.4
Deriving the Function $A(r_1, r_2)$

We can thus derive $A(r_1, r_2)$ if we can derive equations for ψ, ϕ, and $T(r_1, r_2)$. By following the format discussed in the introductory remarks, we can use the Law of Cosines repeatedly to obtain:

$$1 = r_2^2 + r_1^2 - 2r_1 r_2 \cos(\phi) \quad \rightarrow \quad \phi = \cos^{-1}\left(\frac{r_2^2 + r_1^2 - 1}{2r_1 r_2}\right)$$

$$r_2^2 = 1 + r_1^2 - 2r_1 \cos(\pi - \psi) \quad \rightarrow \quad \psi = \cos^{-1}\left(\frac{r_2^2 - 1 - r_1^2}{2r_1}\right)$$

We can thus write $T(r_1, r_2)$ in terms of known quantities to get $T(r_1, r_2) = r_1 r_2 \sin(\phi)/2$. Combining this all together we have:

$$A(r_1, r_2) = \cos^{-1}\left(\frac{r_2^2 - 1 - r_1^2}{2r_1}\right) - r_2^2 \cos^{-1}\left(\frac{r_2^2 + r_1^2 - 1}{2r_1 r_2}\right) + r_1\left(1 - \left(\frac{r_2^2 + r_1^2 - 1}{2r_1 r_2}\right)\right)^{1/2} \tag{B.2}$$

We can thus write:

$$P[R_2 \leq r_2 \mid R_1 = r_1] = 1 - \frac{A(r_1, r_2)}{A(r_1, 1)} \qquad 1 \leq r_2 \leq 1 + r_1$$

The conditional density $h_2(r_2 \mid r_1)$ can thus be found as:

$$h_2(r_2 \mid r_1) = \frac{-A'(r_1, r_2)}{A(r_1, 1)} \qquad 1 \leq r_2 \leq 1 + r_1 \tag{B.3}$$

where after tedious calculation we have:

$$A'(r_1, r_2) = \frac{r_2^2 V_3}{2V_5} + \frac{r_1 V_5}{2} - r_2\left(\frac{1}{2r_1 V_4} + \frac{r_1 V_4 V_2}{2V_5} + \cos^{-1}(V_2)\right)$$

where we have defined:

$$V_1 = \frac{r_2^2 - 1 - r_1^2}{2r_1} \qquad V_2 = \frac{r_2^2 + r_1^2 - 1}{2r_1 r_2}$$

$$V_3 = \frac{r_2^2 + 1 - r_1^2}{2r_1 r_2^2} \qquad V_4 = (1 - V_1^2)^{1/2}$$

$$V_5 = (1 - V_2^2)^{1/2}$$

Using this conditional density and the density for R_1 just found, we can find the unconditional density for R_2 as:

$$h_2(r_2) = \int_{r_2-1}^{1} h_2(r_2 \mid r_1)\, 2r_1\, dr_1 \qquad 1 \leqslant r_2 \leqslant 2 \tag{B.4}$$

B.3 The Density for Θ

We will derive the density for Θ for $\theta \leqslant \pi$ since the case for $\theta > \pi$ is completely analogous. For this derivation we will make use of figure B.5.

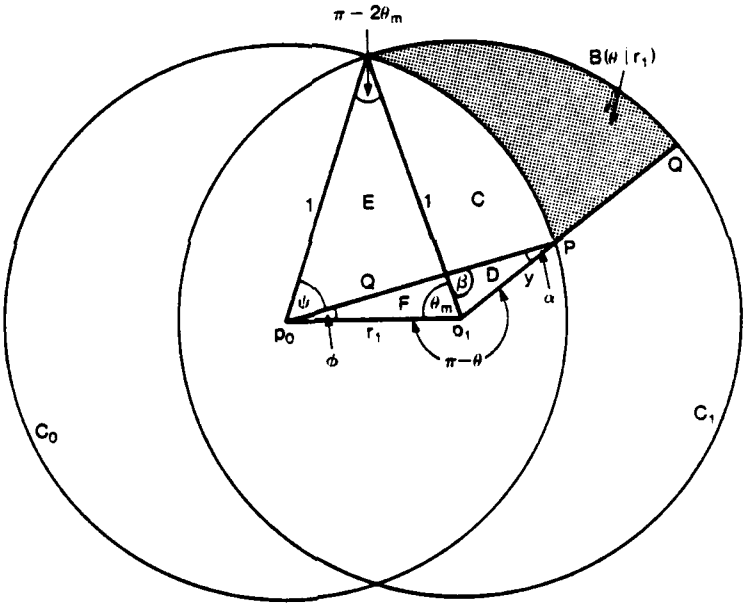


FIGURE B.5
Calculating the density for Θ.

In this figure, $B(\theta \mid r_1)$ is the shaded area, D is the triangle formed by points p_1, and P and angle α, and area C plus D make up the sector of unit radius and angle $\theta - \theta_m$. Also in the figure, F is the triangle of vertices p_0, p_1, and angle θ_m, and E is the triangle made up of ver-

tex p_0 and angles ψ and $\pi - 2\theta_m$. We can write the conditional density for Θ as:

$$P[\Theta \leqslant \theta \mid R_1 = r_1] = \frac{B(\theta \mid r_1)}{A(r_1, 1)} \quad \theta \geqslant \theta_m = \cos^{-1}(r_1/2)$$

From the figure it is clear that:

$$B(\theta \mid r_1) = (\theta - \theta_m)/2 - (C + D)$$

We thus need to derive the areas C and D. First we will derive the length of y. By the Law of Cosines we have:

$$1 = r_1^2 + y^2 - 2r_1 y \cos(\theta)$$

which we can solve by the quadratic formula to obtain:

$$y = r_1 \cos(\theta) + (1 - r_1^2 \sin^2(\theta))^{1/2}$$

Using this we will next derive C and D:

$$\frac{1}{\sin(\theta)} = \frac{r_1}{\sin(\alpha)} \quad \rightarrow \quad \alpha = \sin^{-1}(r_1 \sin(\theta))$$

$$\beta = \pi - \alpha - (\theta - \theta_m) \quad , \quad \psi = 2\theta_m - \beta$$

$$\frac{r_1}{\sin(\pi - \beta)} = \frac{Q}{\sin(\theta_m)} \quad \rightarrow \quad Q = \frac{r_1 \sin(\theta_m)}{\sin(\pi - \beta)}$$

$$E = \tfrac{1}{2} Q \sin(\psi) \quad , \quad \phi = \pi - \alpha - \theta$$

$$F = \tfrac{1}{2} r_1 Q \sin(\phi)$$

and then C and D can be written in terms of known quantities as:

$$C = \frac{\psi}{2} - E \qquad D = \tfrac{1}{2} r_1 y \sin(\theta) - F$$

The conditional density, $k(\theta \mid r_1)$ then is:

$$k(\theta \mid r_1) = \frac{dB(\theta \mid r_1)}{d\theta A(r_1, 1)} \qquad \theta \geqslant \cos^{-1}(r_1/2) \tag{B.5}$$

and the unconditional density can be found as:

$$k(\theta) = \int_{\max(2\cos(\theta), 0)}^{1} k(\theta \mid r_1)\, 2r_1 dr_1$$

While we are at this point, we can easily derive $j(z \mid \theta, r_1)$, the conditional density of the euclidean distance from point p_1 to p_2 given r_1 and θ. In figure B.5 we see that under these conditions, Z must lie on the line from P to Q. Because points are assumed to be distributed according to the Poisson distribution, the conditional density will be uniform over the length of the line segment from P to Q, and thus using the derived expression for y above, we can write:

$$j(z \mid \theta, r_1) = 1/(1-y) \qquad y \leqslant z \leqslant 1 \tag{B.6}$$

B.4 The Density of R_3

The calculation becomes considerably more complex and tedious as we progress to the third hop. As in the derivation for R_2, we will first calculate the distribution, conditioned appropriately, by forming a ratio of two areas. Figure B.6 shows a 3 hop configuration for a given r_1, θ, and z. Any node in the third hop must lie within that portion of C_2 that lies outside C_1 and C_2. In the figure we have defined the angles and lengths that are needed to calculate the area contained in the two wedges $A_u(y \mid r_1, \theta, z)$ and $A_l(y \mid r_1, \theta, z)$ that are highlighted in the figure. If the total area lying only in C_3 is $T(r_1, \theta, z)$, and if $y_m \leqslant y \leqslant 1+r_1$ (bounds to be derived later) then:

$$P[R_3 \leqslant y \mid R_1 = r_1, \Theta = \theta, Z = z] = \frac{A_u(y \mid r_1, \theta, z) + A_l(y \mid r_1, \theta, z)}{T(r_1, \theta, z)} \tag{B.7}$$

The upper bound, $y \leqslant 1+r_1$, arises because at $y = 1+r_1$ the circle of radius y from p_0 becomes tangent to C_1. For $y' > 1+r_1$, as shown in the figure, the area is increased by the circular annulus labeled $B(y' \mid r_1, \theta, z)$. The area for this band was derived in equation (B.2), and thus we have:

$$B(y' \mid r_1, \theta, z) = A(1+r_1, 1+r_1) - A(y', 1+r_1)$$

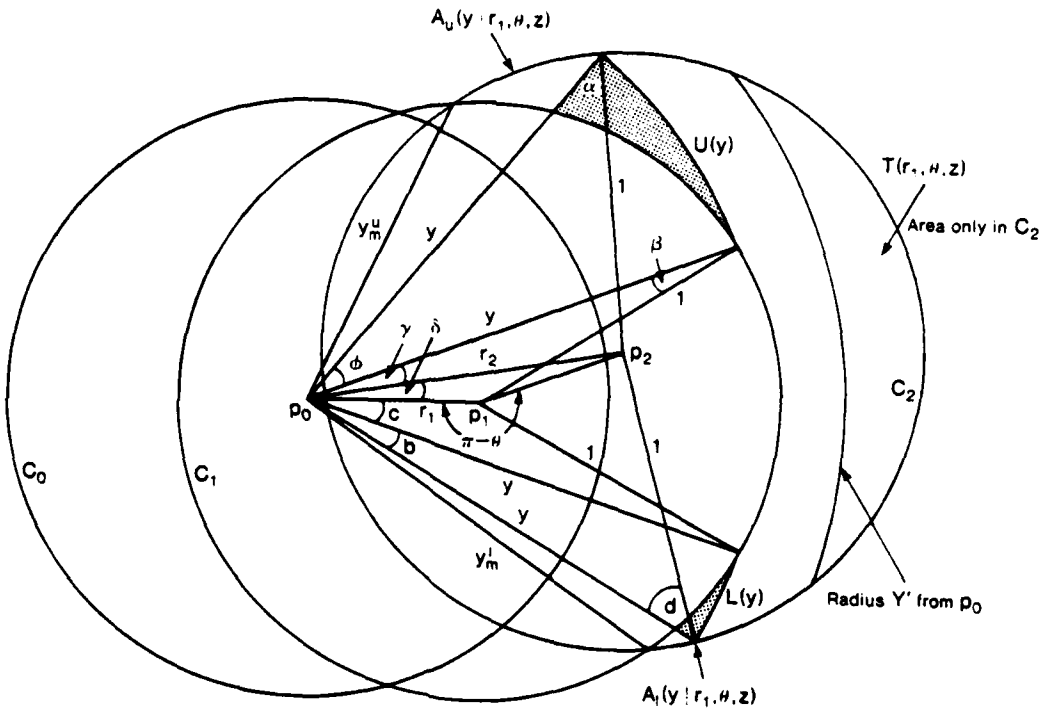


FIGURE B.6
Deriving the third hop density function.

We can thus write:

$$P[R_3 \leqslant y' \mid R_1 = r_1, \Theta = \theta, Z = z] = \frac{A_{l+u}(1+r1 \mid r1, \theta, z) + B(y' \mid r_1, \theta, z)}{T(r_1, \theta, z)} \tag{B.8}$$

where $1+r_1 \leqslant y \leqslant 1+r_2 \leqslant 3$, $A_{l+u}(\cdot) = A_l(\cdot) + A_u(\cdot)$, and the upper bound is at the point where the arc of radius $1+r_2$ from p_0 becomes tangent to C_2, and the expression for r_2 can be found in the previous section. It is clear that $T(r_1, \theta, z)$ can then be expressed as:

$$T(r_1, \theta, z) = A_{l+u}(1+r1 \mid r1, \theta, z) + A(1+r_1, 1+r_1)$$

The only unknown quantities needed to calculate the distribution then are equations for $A_u(y \mid r_1, \theta, z)$ and $A_l(y \mid r_1, \theta, z)$. This can be done using the defined angles and lengths

shown in figure B.6. First we will derive $A_u(y \mid r_1, \theta, z)$.

The objective of this derivation will be to find the length $U(y)$, of the arc subtended by angle ϕ. Once this is found, the integration of this from y_m^u, the minimum allowed value for r_3, to y, will be equal to the area of the shaded wedge. The derivation proceeds by a rather long deduction of angles and lengths:

$$r_2^2 = r_1^2 + z^2 - 2r_1 z\cos(\theta) \quad \rightarrow \quad r_2 = (r_1^2 + z^2 - 2r_1 z\cos(\theta))^{1/2}$$

$$r_2^2 = y^2 + 1 - 2y\cos(\alpha) \quad \rightarrow \quad \alpha = \cos^{-1}\left(\frac{y^2 + 1 - r_2^2}{2y}\right)$$

$$r_1^2 = y^2 + 1 - 2y\cos(\beta) \quad \rightarrow \quad \beta = \cos^{-1}\left(\frac{y^2 + 1 - r_1^2}{2y}\right)$$

$$\frac{z}{\sin(\delta)} = \frac{r_2}{\sin(\theta)} \quad \rightarrow \quad \delta = \sin^{-1}\left(\frac{z\sin(\theta)}{r_2}\right)$$

$$\frac{1}{\sin(\gamma + \delta)} = \frac{r_1}{\sin(\beta)} \quad \rightarrow \quad \gamma = \sin^{-1}\left(\frac{\sin(\beta)}{r_1}\right) - \delta$$

$$\frac{1}{\sin(\phi + \gamma)} = \frac{r_2}{\sin(\alpha)} \quad \rightarrow \quad \phi = \sin^{-1}\left(\frac{\sin(\alpha)}{r_2}\right) - \gamma$$

Thus we have $U(y) = y\phi$ expressed in terms of r_1, θ, and z, and we can write:

$$A_u(y \mid r_1, \theta, z) = \int_{y_m^u}^{y} U(r)\, dr \quad y_m^u \leqslant y \leqslant 1 + r_1$$

We can determine y_m^u from figure B.7, where we see that:

$$y_u = \left(r_1^2 + 1 - 2\, r_1 \cos(\theta - \cos^{-1}(z/2))\right)^{1/2}$$

and thus $y_m^u = \max(r_2, y_u)$.

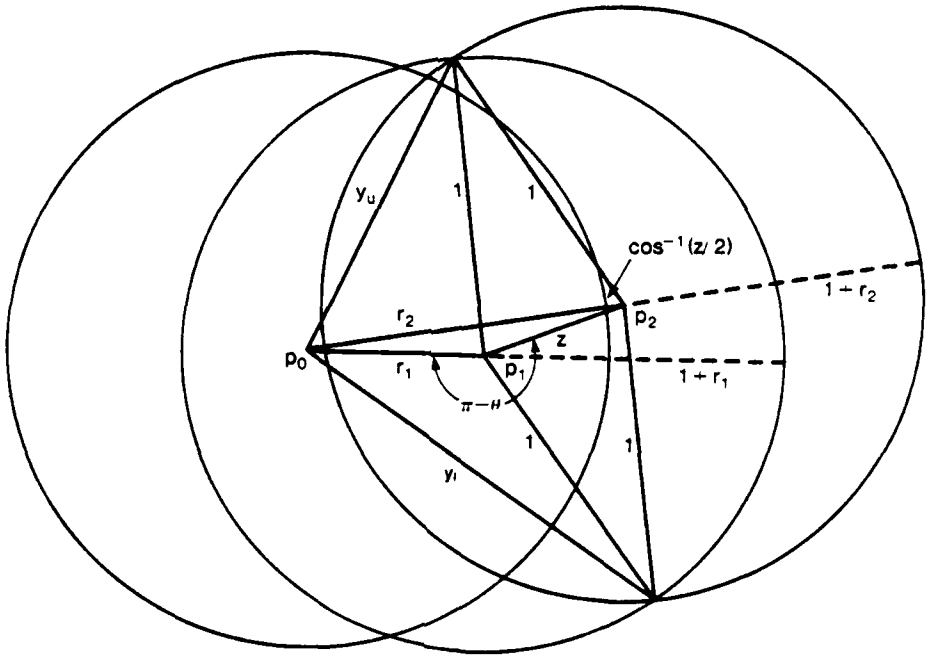


FIGURE B.7
Deriving the upper and lower bounds.

The derivation of $A_l(y \mid r_1, \theta, z)$ proceeds along similar lines, and we have:

$$1 = r_1^2 + y^2 - 2r_1 y\cos(c) \quad \rightarrow \quad c = \cos^{-1}\left(\frac{r_1^2 + y^2 - 1}{2r_1 y}\right)$$

$$r_2^2 = 1 + y^2 - 2y\cos(d) \quad \rightarrow \quad d = \cos^{-1}\left(\frac{y^2 + 1 - r_2^2}{2y}\right)$$

$$\frac{r_2}{\sin(d)} = \frac{1}{\sin(\delta + c + b)} \quad \rightarrow \quad b = \sin^{-1}\left(\frac{\sin(d)}{r_2}\right) - c - \delta$$

and thus $L(y) = by$ is also expressed in terms of r_1, θ, and z. We then have:

$$A_1(y \mid r_1, \theta, z) = \int_{y_m} L(r)\, dr \qquad y_m \leqslant y \leqslant 1 + r_1$$

where again from figure B.7 we have:

$$y_l = \left[r_1^2 + 1 - 2 r_1 \cos(2\pi - \theta - \cos^{-1}(z/2))\right]^{1/2}$$

and thus $y_m^l = \max(r_2, y_l)$.

We can now express the conditional distribution $H(r_3 \mid r_1, \theta, z)$ in terms of r_1, θ, and z. The conditional density for R_3 can then be found by differentiating:

$$h_3(r_3 \mid r_1, \theta, z) = \frac{dH(r_3 \mid r_1, \theta, z)}{dr_3} \tag{B.9}$$

Using equations (B.1), (B.5), (B.6), and (B.9) we can find the unconditional density for R_3 as:

$$h_3(r_3) = \int\int\int h_3(r_3 \mid r_1, \theta, z)\, j(z \mid \theta, r_1)\, k(\theta \mid r_1)\, h_1(r_1)\, dz\, d\theta\, dr_1 \tag{B.10}$$

$$\cos^{-1}(r_1/2) \leqslant \theta \leqslant \pi$$

$$0 \leqslant r_1 \leqslant 1$$

$$1 \leqslant r_3 \leqslant 3$$

APPENDIX C
Derivation of $\Pi(S,x,y)$

In this appendix we will determine a formula for $\Pi(S,x,y)$ and this will permit us to explain the particular definition of the rates that were defined in Chapter 5 (Equations (5.1) and (5.2)). It turns out that with these rate definitions, determining the steady state probability distribution for any given topology is mathematically tractable because the Markovian process defined by these rates define a *reversible process* [King69a]. Intuitively a reversible process $X(t)$ is one in which the direction of time has no effect on the statistics of the process and thus $X(t)$ and $X(-t)$ have the same probability distribution. Finding closed form solutions for reversible processes is simplified because that reversible processes satisfy detailed balanced equations. Let $\Pi(S)$ be the steady state probability of state S and let $q(S,S')$ be the rate at which transitions between states S and S' occur. The detailed balanced equations state:

$$\Pi(S)\ q(S,S') = \Pi(S')\ q(S',S) \tag{C.1}$$

In general Markov processes satisfy global balance equations that equate the total probability flux entering a state to that leaving the state. These equations state that:

$$\sum_{S'} \Pi(S')\ q(S',S) = \sum_{S'} \Pi(S)\ q(S,S') \tag{C.2}$$

We can depict equation (C.1) in Figure C.1 where the equation states that the net probability flux across the cut is zero. For reversible systems, equation (C.2) shows that the net flow across arcs connecting any two states (the cut in Figure C.2) is equal to zero. Equation (C.2) can be used to simplify finding closed form solutions if reversibility can be proven, and a useful tool to do this is *Kolmogorov's criteria.*

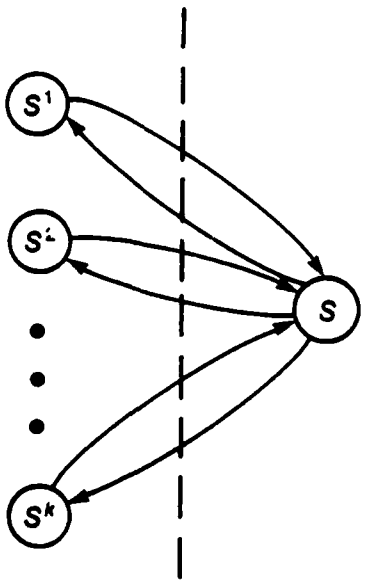


FIGURE C.1
Global Flow Balance Equations.

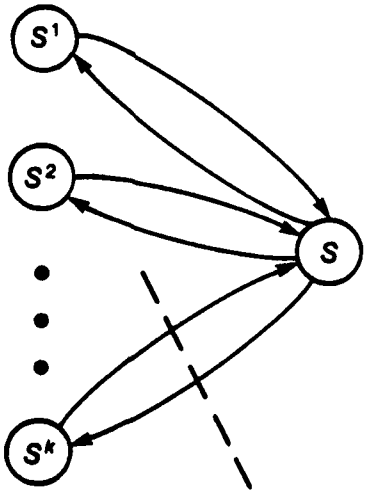


FIGURE C.2
Local Flow Balance Equations.

Theorem: (Kolmogorov's criteria)
Let $S^1, S^2, \cdots, S^k, S^{k+1} = S^1$ be any sequence of states. Then a Markov process defined over these states is reversible if and only if the rate transitions satisfy:

$$q(S^1, S^2)\, q(S^2, S^3) \cdots q(S^{k-1}, S^k)\, q(S^k, B^{k+1}) = q(S^{k+1}, S^k)\, q(S^k, S^{k-1}) \cdots q(S^3, S^2)\, q(S^2, S^1)$$

■

To determine $\Pi(S,x,y)$ for our system, we will first need to make some preliminary definitions. We can imagine that the network for any particular state S, is a labeled graph $G(S)$ in which the edges are determined by the adjacency sets A_i and the state vector S determines the labeling. We will define an *i-arc* (i=0 or i=1) to be an arc in $G(S)$ which has both nodes labeled i. The number of i-arcs in $G(S)$ will be denoted as $B_i(S)$ and $N^j_i(S)$ will be the number of neighbors of node j that are labeled i. Let $T_i(S)$ be an operator acting on state $S = (s_1(S), s_2(S), \cdots, s_n(S))$ which complements the i^{th} component, thus:

$$T_i(S) = (s_1(S), s_2(S), \cdots, s_{i-1}(S), \bar{s}_i(S), s_{i+1}(S), \cdots, s_n(S))$$

and we will say a transition is a 0⟶1 transition if $s_i(S) = 0$ and for some j, we have $T_j(S) = S'$ (1⟶0 transitions are define in the analogous way). We can now prove a lemma that will be used to show that the process defined by rate equations (5.1) and (5.2) is reversible.

Lemma: For any state S, $N^j_i(S) = | B_i(S) - B_i(T_j(S)) |$

Proof: Let $B'_i(S)$ equal the number of i-arcs in the sub-graph consisting of all nodes not contained in $A_j \bigcup \{j\}$. It is clear that $B'_i(S) = B'_i(T_j(S))$. Let k be the number of i-arcs (and thus the number of neighbors labeled i) of which node j is a member in state S and k' be the number in state $T_j(S)$. Since $T_j(S)$ and S differ in the j^{th} component exactly one of k and k'

will be non-zero. We can write:

$$\begin{aligned} B_j(S) &= k + B'_j(S) \\ B_j(T_j(S)) &= k' + B'_j(S) \end{aligned}$$

and thus

$$| B_j(S) - B_j(T_j(S)) | = | k - k' |$$

■

Observe that for a known 0—1 transition involving node j we can eliminate the absolute value sign in the above to get $N_0^j(S) = B_0(S) - B_0(T_j(S))$. In a like manner, if node j is involved in a 1—0 transition we have $N_1^j(S) = B_1(T_j(S)) - B_1(S)$. We are now in a position to prove that the Markov process described before is reversible.

Theorem: The Markov process described by rate definitions (5.1) and (5.2) defines a reversible Markov process.

Proof: We will show that Kolmogorov's criteria is satisfied. For any closed sequence of states $S^1, S^2, \cdots, S^l, \ldots$, Kolmogorov's criteria is trivially satisfied if $S^{l+1} \neq T_j(S^l)$ for some j since the probability of two or more components of S^l and S^{l+1} differing is zero. Thus assume that the sequence of states consists of single component transitions. Since there are only two types of transitions, all closed cycles must contain an even number of states. Let $S^1, S^2, \cdots, S^{2m}, S^{2m+1} = S^1$ be such a sequence. Let

$$\Psi_1 = q(S^1, S^2)\, q(S^2, S^3) \cdots q(S^{2m}, S^{2m+1})$$

$$\Psi_2 = q(S^{2m+1}, S^{2m})\, q(S^3, S^2)\, q(S^2, S^1)$$

We must show that $\Psi_1 = \Psi_2$ and will refer to their corresponding sequences as the forward and backward sequence respectively. We first note that since there are an equal number of 0—1 and 1—0 transitions in Ψ_1 and Ψ_2 we can write $\Psi_i = \mu^m \gamma_0^m \Phi_i$ where Φ_i contains all the fac-

tors of x and y. We thus must show that $\Phi_1 = \Phi_2$. Concentrating first on the exponent of x in these expressions, define the two sets:

$$C_1 = \{ (S^i, S^{i-1}) \mid \exists j \; T_j(S^i) = S^{i-1}, \; s_j(S^i) = 0 \}$$

$$C_2 = \{ (S^{i+1}, S^i) \mid \exists j \; T_j(S^{i+1}) = S^i, \; s_j(S^{i-1}) = 0 \}$$

In words, C_1 contains the 0→1 transitions for the forward sequence and C_2 for the backward sequence. Since a 0→1 transition in the forward sequence is a 1→0 transition in the backward sequence, we know that $C_1 \bigcap C_2 = \varnothing$. Since C_1 only contains 0→1 transitions we can write, using the lemma, the exponent of x in the forward E_1, and backward E_2, sequences as:

$$E_i = \sum_{(S, S') \in C_i} B_0(S) - B_0(S')$$

Suppose now that $S^i, S^{i-1}, \cdots S^{i+k}$ is a sub-sequence of states from the forward sequence such that $(S^i, S^{i+1}) \in C_1$. Since the portion of E_1 for this sequence alternates sign, the sum telescopes and we can write for this sub-section:

$$\sum_{j=i}^{i=i+k-1} B_0(S^j) - B_0(S^{j+1}) = B_0(S^i) - B_0(S^{i+k})$$

and since for this sequence of states, $(S^{i+1}, S^i) \notin C_2 \; j = i, i+1, \cdots, i+k-1$, we can conclude that E_2 is not affected by this telescoping. It is thus clear that in calculating E_1 and E_2 we only have to look at sections of either sequence where a change in the type of state transition occurs. Denoting these places of state transition as S'^i we have:

$$E_1 = [B_0(S'^1) - B_0(S'^2)] + [B_0(S'^3) - B_0(S'^4)] + \cdots + [B_0(S'^{2k-1}) - B_0(S'^{2k})]$$

$$E_2 = [B_0(S'^1) - B_0(S'^{2k})] + [B_0(S'^{2k-1}) - B_0(S'^{2k-2})] + \cdots + [B_0(S'^3) - B_0(S'^2)]$$

and thus see that $E_1 = E_2$. An analogous argument can be made for the exponent of y concluding the proof.

■

Knowing that the Markov process is reversible allows us to use the detailed balance equations (5.1) to prove the following theorem.

Theorem C.1: The equilibrium probability for state S with transitions rates defined by equations (5.1) and (5.2) is given by:

$$\Pi(S,x,y) = C\,\rho^{M(S)}\,x^{-B_0(S)}\,y^{B_1(S)}$$

where:

$$M(S) = \sum_{i=1}^{n} s_i(S)$$

$$C = \Pi(0,x,y)\,x^{-B_0(0)}$$

$$\rho = \gamma_0/\mu$$

$$\Pi(0,x,y) = \left[x^{-B_0(0)} \sum_{S} \rho^{M(S)}\,x^{-B_0(S)}\,y^{B_1(S)}\right]^{-1}$$

Proof: To avoid cumbersome notation let $\Pi(S,x,y) = \Pi(S)$. Since the process is reversible we can use the detailed balance equations to state:

$$\frac{\Pi(S)}{\Pi(T_i(S))} = \frac{q(T_i(S),S)}{q(S,T_i(S))} \tag{C.3}$$

If $s_i(S) = 1$ we can write equation (C.3) as:

$$\frac{\Pi(S)}{\Pi(T_i(S))} = \rho\,x^{N_0(S)}\,y^{N_1(S)}$$

which by using the lemma can be re-written (for the case of a 0→1 transition) as:

$$\frac{\Pi(S)}{\Pi(T_i(S))} = \rho\,x^{B_0(T_i(S))-B_0(S)}\,y^{B_1(S)-B_1(T_i(S))} \qquad s_i(S)=1 \tag{C.4}$$

We can use this relationship to write $\Pi(S)$ in terms of $\Pi(0)$ by telescoping a product of rate

ratios. Suppose $i_1, i_2, i_3, \cdots, i_{M(S)}$ are the indices of S which are equal to 1. Define the following operator:

$$F_0 = 1$$
$$F_j = T_{i_j} F_{j-1} \qquad j = 1, 2, \cdots, M(S)$$

Observe that $F_0(S) = S$ and $F_{M(S)}(S) = 0$ (a vector of all zeros). We can then write equation (C.3) as:

$$\frac{\Pi(S)}{\Pi(0)} = \frac{\Pi(F_0(S))}{\Pi(F_1(S))} \frac{\Pi(F_1(S))}{\Pi(F_2(S))} \cdots \frac{\Pi(F_{M(S)-1}(S))}{\Pi(F_{M(S)}(S))}$$
$$= \frac{q(F_1(S), F_0(S))}{q(F_0(S), F_1(S))} \frac{q(F_2(S), F_1(S))}{q(F_1(S), F_2(S))} \cdots \frac{q(F_{M(S)}(S), F_{M(S)-1}(S))}{q(F_{M(S)-1}(S), F_{M(S)}(S))}$$

Which can be simplified by using equation (C.4) to:

$$\frac{\Pi(S)}{\Pi(0)} = \rho^{M(S)} x^{\sum_{i=0}^{M(S)-1} B_0(F_{i-1}(S)) - B_0(F_i(S))} y^{\sum_{i=0}^{M(S)-1} B_1(F_i(S)) - B_1(F_{i-1}(S))}$$

Once again these sums telescope and we are left with:

$$\frac{\Pi(S)}{\Pi(0)} = \rho^{M(S)} x^{B_0(F_{M(S)}(S)) - B_0(F_0(S))} y^{B_1(F_0(S)) - B_1(F_{M(S)}(S))}$$

We can thus write:

$$\Pi(S) = \Pi(0)\, \rho^{M(S)}\, x^{B_0(0) - B_0(S)}\, y^{B_1(S) - B_1(0)}$$

and using the fact that $B_1(0) = 0$ and defining $C = \Pi(0)\, x^{-B_0(0)}$ we finally obtain:

$$\Pi(S) = C\, \rho^{M(S)}\, x^{-B_0(S)}\, y^{B_1(S)} \tag{C.5}$$

We can determine $\Pi(0)$ by using the normalization constraint $\sum_S \Pi(S) = 1$ to determine the equation for $\Pi(0)$ given in the statement of Theorem C.1.

■

References

[Abra70a] N. Abramson, "The ALOHA System — Another Alternative for Computer Communications," *AFIPS Conference Proceedings, FJCC* **37**, pp.281-285 (1970).

[Abra73a] N. Abramson, "Packet Switching with Satellites," *AFIPS Conference Proceedings, NCC* **42**, pp.695-703 (June 1973).

[Abra77a] N. Abramson, "The Throughput of Packet Broadcasting Channels," *IEEE Transactions on Communications* **COM-25**(1), pp.117-128 (January 1977).

[Akav78a] G. Y. Akavia, *Hierarchical Organizations of Distributed Packet-Switching Communication Systems*, Computer Science Department, University of California, Los Angeles (March 1978). Ph.D. Dissertation.

[Avri76a] M. Avriel, *Nonlinear Programming: Analysis and Methods*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey (1976).

[Cant74a] D. G. Cantor and M. Gerla, "Capacity Allocation in Distributed Computer Networks," *Proceedings of the Seventh Hawaii International Conference on Systems Sciences*, pp.115-117, University of Hawaii (January 8-10 1974).

[Davi80a] D. H. Davis and S. A. Gronemeyer, "Performance of Slotted Aloha Random Access with Delay Capture and Randomized Time of Arrival," *IEEE Transactions on Communications* **COM-28**(5), pp.703-710 (May 1980).

[Fran71a] H. Frank, I. T. Frisch, W. Chou, and R. Van Slyke, "Optimal Design of Centralized Computer Networks," *Networks* **1**(1), pp.43-57, Wiley (1971).

[Frat73a] L. Fratta, M. Gerla, and L. Kleinrock, "The Flow Deviation Method: An Approach to Store-and-Forward Communication Network Design," *Networks* **3**, pp.97-133 (1973).

[Frat80a] L. Fratta and D. Sant, "Some Models of Packet Radio Networks with Capture," *Proceedings of the Fifth International Conference on Computer Communication*, pp.155-161 (October 27-30, 1980).

[Howa60a] R. A. Howard, *Dynamic Programming and Markov Processes*, M.I.T. Press, Cambridge, Mass. (1960).

[Kahn78a] R. E. Kahn, S. A. Gronemeyer, J. Burchfiel, and R. C. Kunzelman, "Advances in Packet Radio Technology," *Proceedings of the IEEE* **66**, pp.1468-1496 (November 1978).

[Kell79a] F. P. Kelly, *Reversibility and Stochastic Networks*, John Wiley and Sons (1979).

[Kend63a] M. Kendall and P. Moran, *Geometrical Probability*, Griffin, London (1963).

[King69a] J. F. C. Kingman, "Markov Population Processes," *J. Appl. Prob.* **6** (1969).

[Klei64a] L. Kleinrock, *Communication Nets: Stochastic Message Flow and Delay*, McGraw-Hill, New York (1964). Out of print. Reprinted by Dover Publications, 1972.

[Klei75a] L. Kleinrock and F. A. Tobagi, "Packet Switching in Radio Channels: Part I — Carrier Sense Multiple-Access Modes and Their Throughput-Delay Characteristics," *IEEE Transactions on Communications* **COM-23**(12), pp.1400-1416 (December 1975).

[Klei76a] L. Kleinrock, *Queueing Systems. Vol II., Computer Applications*, Wiley-Interscience, New York (1976).

[Klei78a] L. Kleinrock and J. Silvester, "Optimum Transmission Radii for Packet Radio Networks or Why Six is a Magic Number," *Conference Record. National Telecommunications Conference*, pp.4.3.1-4.3.5 (December 1978).

[Lam75a] S. S. Lam and L. Kleinrock, "Dynamic Control Schemes for a Packet Switched Multi-Access Broadcast Channel," *AFIPS Conference Proceedings, NCC* **44**, pp.143-154 (May 1975).

[Lick78a] J. C. R. Licklider and A. Vezza, "Applications of Information Networks," *Proceedings of the IEEE* **66**(11) (November 1978).

[Litt61a] J. Little, "A Proof of the Queueing Formula $L = \lambda W$," *Operations Research* **9**(2), pp.383-387 (March 1961).

[Liu80a] J. Liu, "A Distributed Routing Algorithm in Mobile Packet Radio Networks," Technical Report, University of Illinois, Urbana, Ill. (1980).

[Meis71a] B. Meister, H. Muller, and H. Rudin, "New Optimization Criteria for Message-Switching Networks," *IEEE Transactions on Communication Technology* **COM-19**, pp.256-260 (June 1971).

[Mine70a] H. Mine and S. Osaki, *Markovian Decision Processes*, American Elsevier (1970).

[Moll81a] M. L. Molle, "Unifications and Extensions of the Multiple Access Communications Problem," CSD Report No. 810730 (UCLA-ENG-8118), Computer Science Department, University of California, Los Angeles (July 1981). Ph.D. Dissertation.

[Robe72b] L. G. Roberts, "ALOHA Packet System With and Without Slots and Capture," ASS Note 8 (NIC 11290), ARPA Network Information Center, Stanford Res. Inst., Menlo Park, Ca. (June 1972). reprinted in *Computer Communication Review*, Vol. 5, pp. 28-42 (April 1975)

[Robe72a] L. G. Roberts, "Extensions of Packet Communication Technology to a Hand Held Personal Terminal," *AFIPS Conference Proceedings, SJCC* **40**, pp.295-298 (1972).

[Silv80a] J. Silvester, "On Spatial Capacity of Packet Radio Networks," UCLA-ENG-8021, Computer Science Department, University of California, Los Angeles (May 1980). Ph.D. Dissertation.

[Toba74a] F. A. Tobagi, "Random Access Techniques for Data Transmission over Packet Switched Radio Networks," UCLA-ENG-7499, Computer Science Department, University of California, Los Angeles (December 1974). Ph.D. Dissertation.

[Toba75a] F. A. Tobagi and L. Kleinrock, "Packet Switching in Radio Channels: Part II. — The Hidden Terminal Problem in Carrier Sense Multiple-Access and the Busy Tone Solution," *IEEE Transactions on Communications* **COM-23**(12), pp.1417-1433 (December 1975).

[Toba80a] F. A. Tobagi, "Analysis of a Two-Hop Centralized Packet Radio Network-Part I: Slotted Aloha," *IEEE Transactions on communications* **Com-28**(2), pp.196-207 (February 1980).

[Toba80b] F. A. Tobagi, "Analysis of a Two-Hop Centralized Packet Radio Network-Part II: Carrier Sense Multiple Access," *IEEE Transactions on Communications* **Com-28**(2), pp.208-216 (February 1980).

[Toth64a] L. Fejes Toth, *Regular Figures*, Oxford: Pergamon Press (1964).

[Will79a] R. Williams, *The Geometrical Foundation of Natural Structure*, Dover Publications, Inc. (1979).

END

FILMED

11-83

DTIC